DE GRUYTER

LEXICOGRAPHY OF CORONAVIRUS-RELATED NEOLOGISMS

Edited by Annette Klosa-Kückelhaus and Ilan Kernerman

LEXICOGRAPHICA. SERIES MAIOR

DE G

Lexicography of Coronavirus-related Neologisms

LEXICOGRAPHICA

Series Maior

Supplementary Volumes to the International Annual for Lexicography
Suppléments à la Revue Internationale de Lexicographie
Supplementbände zum Internationalen Jahrbuch für Lexikographie

Edited by
Rufus Hjalmar Gouws, Ulrich Heid, Thomas Herbst,
Anja Lobenstein-Reichmann, Oskar Reichmann,
Stefan J. Schierholz and Wolfgang Schweickard

Volume 163

Lexicography of Coronavirus-related Neologisms

Edited by
Annette Klosa-Kückelhaus and Ilan Kernerman

DE GRUYTER

This publication has received funding from Leibniz-Gemeinschaft (Publikationsfonds für Monografien) and Leibniz-Institut für Deutsche Sprache (Mannheim). / Diese Publikation wurde gefördert durch den Publikationsfonds für Monografien der Leibniz-Gemeinschaft und durch das Leibniz-Institut für Deutsche Sprache (Mannheim).

ISBN 978-3-11-079556-1
e-ISBN (PDF) 978-3-11-079808-1
e-ISBN (EPUB) 978-3-11-079831-9
ISSN 0175-9264
DOI https://doi.org/10.1515/9783110798081

Library of Congress Control Number: 2022945605

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Contents

Annette Klosa-Kückelhaus, Ilan Kernerman

Lexicography of Coronavirus-related neologisms: An introduction

1 Background

This volume of *Lexicographica. Series Maior* focuses on lexicographic neology and neological lexicography concerning COVID-19 neologisms, featuring papers originally presented at the third Globalex Workshop on Lexicography and Neology (GWLN 2021[1]). GWLN 2021 was held online in conjunction with Australex 2021,[2] with a focus on neologisms arising in relation to the COVID-19 pandemic. Papers discussing various issues related to the detection of such neologisms – including new words, new meanings of existing words, and new multiword units – and their representation in lexicography and dictionaries were invited to offer cross-world views on lexicographic detection and representation of Coronavirus-driven neologisms for different languages. Similar challenges regarding COVID-19 neologisms and lexicography arise for any contemporary language, for example how to detect such neologisms (corpus analysis and editorial means of identification, evaluation of other data, e.g. blogs and chats) or how dictionary users can help with finding and informing about them. But also the extent of borrowing COVID-19 neologisms from other languages (and which ones), in contrast to the use of word formation processes to enlarge the Coronavirus-related vocabulary in a specific language, needs to be examined, and questions of prescriptive vs. descriptive lexicographic information on such neologisms need to be addressed.

The GWLN series began as a single event conjugated with the 22nd Biennial Meeting of the Dictionary Society of North America (DSNA) at Indiana University, Bloomington, in 2019[3] and included thirteen invited papers from around the world, of which eight formed a special issue of the DSNA's journal *Dictionaries*, published the following year (2020, 41.1[4]). GWLN-2[5] was planned in conjunction with the Euralex 2020 Congress (Alexandropoulos, Greece), but due to the COVID-19 pandemic

1 https://globalex2021.globalex.link/ (last access: 10 June 2022).
2 https://www.adelaide.edu.au/australex/ (last access: 10 June 2022).
3 https://dictionarysociety.com/ (last access: 10 June 2022).
4 https://dictionarysociety.com/wp-content/uploads/2020/05/Dictionaries-41.1-Table-of-Contents.pdf (last access: 10 June 2022).
5 https://globalex2020.globalex.link/gw-euralex2020 (last access: 10 June 2022).

Annette Klosa-Kückelhaus, Leibniz-Institut für Deutsche Sprache, e-mail: klosa@ids-mannheim.de
Ilan Kernerman, K Dictionaries, e-mail: ilan@kdictionaries.com

https://doi.org/10.1515/9783110798081-001

it was partially held online (November 2020)[6] and as a special session at Euralex 2020 online (in 2021),[7] with selected papers published as a special issue of *International Journal of Lexicography* (Klosa-Kückelhaus/Kernerman 2021).[8]

Lexicography has been around for thousands of years and has always had to adapt to developments in society and language, apparently more than ever in the last generation with its increasingly rapid and radical technological changes. Neology has been there forever, driving language from the start and so-to-speak inciting lexicography. Likewise, in recent decades neology has been drawing more attention in research communities and inspiring new practical applications, such as at university or national language observatories or in the language technology industry, as well as with the general public. The speed of novelty in daily life accelerates and the volume of innovations grows exponentially – all defined by language as well as affected by and affecting language. Altogether, there is greater interest in neologisms and in the role of lexicographic resources to capture and disseminate them to the world.

The overall aim of GWLN and its corresponding publications is to explore this intersection of neology and lexicography worldwide, uncover the common factors and highlight individual features, expose and share the findings with each other and enhance mutual understanding, professional competence, and user satisfaction. The main issues in question begin with the identification of neologisms and go on to comprise their categorization and lexicographic treatment and representation. As such, the description in our introduction to the special issue of *Dictionaries* (Klosa-Kückelhaus/Kernerman 2020) is appropriate here, too, and we reproduce it with slight adjustments:

"Neology constitutes a natural, dynamic and multilateral part of all living human languages, whether as a reflection or for facilitation of linguistic communication, and lexicographic interest in neologisms is at least as old as dictionaries themselves. There is a vast field of research of neologisms, pertaining to their origin (stemming from the given language as in new word formation, or loan words from other languages including the dominance of English today, as well as combining both), distribution (in general language and in domain-specific language, that is terminology), identification (applying corpus linguistics methods, editorial methods, user generated candidates, and comparison of different methods), evaluation (such as in blogs and chats), and more. The general definition of neologisms as applied here refers to new words, new multiword units, new elements of word formation, and new meanings of either of them, and addresses lexicography-driven or -oriented aspects, including:

6 For the program, see https://globalex2020.globalex.link/globalex2020-online/ (last access: 10 June 2022).

7 https://euralex2020.gr/ (last access: 10 June 2022).

8 https://academic.oup.com/ijl/issue/34/3 (last access: 10 June 2022).

- How to interoperate lexicographic datasets with online resources and incorporate neologisms into dictionaries (the media, formatting, labelling, etc.)
- How to deal with grammatical/orthographic/pronunciation variation (descriptive vs. prescriptive approaches)
- How to explain meaning with/without encyclopaedic information, and how to use illustrations and audio-visual media
- How well are neologisms that are integrated in dictionaries accepted by the community (issues of rejection of new words and language purism)
- How differently, if at all, should neologisms be treated in different dictionary types (e.g. in historical comprehensive ones as opposed to those focusing on current usage; in monolingual vs. bilingual dictionaries; in special dictionaries of neologisms; in special domain dictionaries)
- How to deal with neologisms that are no longer *new* and with those no longer used
- How can dictionary users help with finding and informing about neologisms

The papers in this volume pursue the discussion on some of these aspects, presenting state-of-the-art research into neology [specific to the COVID-19 pandemic] and ideas on modern lexicographic treatment of neologisms in various dictionary types."

2 This volume

The thirteen papers in this volume focus on ten languages: one Altaic (Korean), one Finno-Ugric (Hungarian), two Germanic (English and German), four Romance (French, Italian, [Brazilian and European] Portuguese and [Pan-American and European] Spanish), and one Slavic (Croatian), as well as the Sign Language of New Zealand. Specialized dictionaries of neologisms are discussed as well as general language ones, monolingual, bilingual and multilingual lexical resources, print and electronic dictionaries. Questions regarding terminology as well as general language and standard and norm regarding COVID-19 neologisms are raised and different methods of detecting candidates in media corpora, as well as by user contributions, are discussed.

The papers are broadly arranged in four groups of three (and four) papers each. The first group features papers regarding English, German, and Korean, respectively, evolving from systemic neological and lexicographic research carried out in their authors' institutions for some years, which conveys solid support and wide perspectives to their findings. The second consists of three papers regarding Spanish neologisms in traditional and upcoming lexicographic contexts from Europe and Latin America. The third presents work on Croatian, Hungarian, Italian, and Portuguese in Portugal and Brazil, i.e. to some extent lesser used languages, which is no less pertinent as for dealing with similar issues. The fourth group of papers extends beyond mainstream lexicography to study COVID-19 neology in relation to

collaborative editing in Wiktionary, to terminology, and to New Zealand sign language. Together, this collection offers rich insights that sometimes overlap while remaining unique.

In **The *Oxford English Dictionary* and the language of Covid-19**, Danica Salazar and Kate Wild offer insight into how the editorial team working on this renowned historical dictionary of English reacted to the challenges posed by the rapid expansion of new vocabulary during the Coronavirus pandemic: "The lexical adaptation necessitated by this global health crisis has been unprecedented in speed and scope, and in response, the *Oxford English Dictionary* (OED) has continually revised its coverage, publishing special updates of Covid-19-related words in 2020 outside of its usual quarterly publication cycle." The *Oxford Languages' monitor corpus of English* and other text databases were used to monitor the development of pandemic-related words, and the authors describe how new lexemes (most prominently *COVID-19*) and words with new meanings (e.g. *bubble*) or new significance (e.g. *social distancing*) were detected and treated lexicographically, for example by revising existing entries and adding new ones. Questions of how the use of terminology in general discourse and regional variation should be transferred into lexicographic information are discussed as well.

Finally, the authors explain how their work expanded beyond the dictionary itself: "The OED's efforts to document the lexical change brought to the English language by the coronavirus pandemic continued throughout 2020, culminating with the *Words of an Unprecedented Year* report, which was published at the end of the year in place of the usual selection of a single Word of the Year (Oxford Languages 2020). This expansive report on the words that defined 2020 features an entire section dedicated to the language of Covid-19." Many dictionary projects around the world reacted in a similar manner and started to publish texts on the COVID-19 vocabulary addressed to the public, as other papers in this volume show.

While the OED as a comprehensive dictionary on general language will only include some highly frequent new lexemes or new meanings into its content, specialized dictionaries on neologisms can be more generous when it comes to the number of new entries. In the paper titled **German Corona-related neologisms and their lexicographic representation**, Annette Klosa-Kückelhaus discusses this question and contrasts two different perspectives: "There are some (neologism) dictionaries that only record neologisms retrospectively, that is after their lexicalization. [. . .] Other neologism dictionaries [. . .] record neologisms [. . .] before they are fully lexicalized, but are nevertheless accepted parts of the lexicon." Presenting data on an online neologism dictionary published by the Leibniz Institute for the German Language (IDS), the author demonstrates how both approaches are combined in one project so that dictionary users may find information on COVID-19 neologisms (new lexemes, new meanings, and new usages) as soon as possible throughout the pandemic development. She also discusses how to detect candidates for inclusion, for example

by continuously evaluating user contributions via a word proposal form on the dictionary webpage.

Overall, this dictionary project seems to have profited from the challenges posed by the rapid vocabulary expansion throughout the pandemic, as "the general awareness for the lexicographic work at IDS and for the usefulness of reliable, up-to-date dictionaries" was raised, "making it worthwhile to immerse in lexicography 'at the pulse of time'."

The emergence and spread of Korean COVID-19 neologisms in news articles and user comments and their lexicographic description is the topic of the paper by Kilim Nam, Jinsan An and Hae-Yung Jung, in which they examine the occurrence frequencies and usage trends of COVID-19 neologisms in news articles and user comments related to the pandemic, to provide information on Korean neologism usage across genre. As "COVID-19 neologisms, in particular, have proliferated for the past year or so, to express, describe, and comment on a global phenomenon, constituting an unprecedented case of profuse and multifaceted neological creativity centered on a single topic", they lend themselves especially well to analyze the differences in distribution and trends across genres. By carrying out secondary collocate and n-gram analyses in addition to frequency and primary collocate analyses, the authors collect data providing a better understanding of the use context for neologisms, in a case study of the neologism *K-quarantine*. Finally, they propose a microstructural model for COVID-19 neologisms that integrates the findings of the study, taking the neologisms *Wuhan Pneumonia* and *K-quarantine* as examples.

The results presented in this paper show that comment data prove invaluable for lexicographic description of neologisms: "The value of comment data in lexicographic description ultimately lies in the pragmatic information and the socio-cultural background it provides on headwords and which are not easily seen in existing dictionaries. Moreover, unlike articles, comments are produced by a multitude of commenters and reflect their emotions and stances in relation to the relevant neologisms, providing dictionary users and future generations with fresh, raw examples of real-life language for neologism headwords." The authors concede, though, that experts need to decide to what extent the politically incorrectness of commenters' language may be used in dictionaries.

Moving to the lexicographic description of COVID-19 neologisms in Spanish, it becomes evident that the question of a lexeme losing its neologism status by being included in general dictionaries also needs to be discussed. In their paper **Lexicographic detection and representation of Spanish neologisms in the COVID-19 pandemic**, Pedro J. Bueno and Judit Freixa "address the neological process and [. . .] reflect on the various stages of it, from the time a neologism is born until the moment it ceases to be one because it has been dictionarised" (i.e. incorporated into a dictionary). Based on their definition of "pandemic neologisms" and the neological process, the authors give information on their corpus data and data analysis methods before presenting three different groups of COVID-19 neologisms: "non-dictionarisable

neologisms", "neologisms in the antechamber of dictionarization", and "dictionarisable neologisms". They also discuss how some of the neologisms found in their study have recently been added to the *Diccionario de la lengua española* (DLE), the authoritative Spanish language dictionary published by the Royal Spanish Academy with participation of the Association of Academies of the Spanish Language**.**

The authors point out that the inclusion of neologisms in dictionaries is "dual property acting on a two-fold plane: that of consolidation in use on the one hand, and that of the criteria governing the elaboration of dictionaries on the other". Some thoughts also go into the different categories that neologisms fall in: short-lived, fleeting ones, and those staying on and becoming fully lexicalized and accordingly recorded in general language dictionaries.

Andreína Adelstein and Victoria de los Ángeles Boschiroli take this discussion a step further in their paper **Spanish neologisms during the COVID-19 pandemic: Changing criteria for their inclusion and representation in dictionaries**, by looking not only into the inclusion of COVID-19 neologisms in (synchronic and historical) general language dictionaries of Spanish, but also into a bilingual English-Spanish dictionary and a Spanish neologism dictionary aiming to cover geolectal variants in six Spanish-speaking countries in Pan-America. The authors describe the different criteria used in the process of inclusion and treatment of the lexemes in those dictionaries starting their study with data obtained by the *Antenas Neológicas Network*, which are "collected exclusively from the written press of the six countries that make up the network". They concede that "this may be regarded as a limitation in terms of diaphasic variation in relation to pandemic vocabulary, but on the other hand, it guarantees a certain degree of institutionalization, which is an essential aspect when considering the inclusion of new words in a general language dictionary."

By comparing how different types of dictionaries include/exclude COVID-19 neologisms, they find that there is "a certain degree of overlap of some features which are traditionally thought to be specific to each type of dictionary: [. . .] Dictionaries which, unlike dictionaries of neologisms (which make no claim to finality of stability regarding the place in the language of the items collected), are not restricted to these phenomena or not supposed to collect them, ended up recording ephemeral or witness items, with a very low or null frequency of use."

In a third perspective on the Spanish language, Magdalena Coll and Mario Barité focus on the inclusion of technical COVID-19 neologisms into a general language dictionary of Spanish in their paper **Specialized voices in the 23rd edition of the *Diccionario de la lengua española*: Analysis of the COVID-19 field and its neologisms**. By analyzing the lexicographic treatment of specialized language neologisms as well as new words beginning with CORONA-, they assess the particularities of the dictionaries in question regarding the incorporation of the new words, as well as the degree of correspondence or complementarity between the last two editions of DLE. The authors demonstrate how "the new additions open up a debate

on the treatment of neologisms in academic lexicography, in a particularly unique scenario".

Here again, the rapid vocabulary expansion and its subsequent lexicographic treatment throughout the COVID-19 pandemic is seen as an "opportunity for lexicography and terminology researchers", who should "discuss and propose consistent solutions for the incorporation of scientific and specialized words into DLE and other Spanish dictionaries" and "leave behind vague criteria for incorporating or excluding scientific terms, scientific definitions not easily understood by a regular audience, conceptual inaccuracies, and somewhat erratic assignments of thematic labels".

In the first paper of the next group in this volume, **How the COVID-19 pandemic is changing the Hungarian language: Building a domain-specific Hungarian/Italian/English dictionary of the COVID-19 pandemic**, Judit Papp looks into ways of compiling a trilingual online dictionary with COVID-19 neologisms using different corpus and dictionary writing tools: "With the creation of the dictionary, my aim is to fill [. . . the] lexicographic gap primarily concerning the Hungarian-Italian language pair and to organize this content in a free online tool (a rich database) that is easy to search and useful for linguists and translators. The third language is English, as the comparison with it is inevitable. [. . .] papers, findings, and results of scientists' experiments relating to COVID-19 are published in English and this means that English plays an important role in the creation of neologisms. In both Hungarian and Italian, we record a certain number of loans, calques, and adaptations".

Here, again, the author interprets the high number of COVID-19 neologisms as a sign for the creativity and vitality of a language (namely Hungarian), and discusses how these aspects affect the lexicographic description in an online dictionary (here a trilingual dictionary of equivalents).

Questions of standardization not only arise regarding terminology, but also in connection with general language. In their paper **Coronavirus-related neologisms: A challenge for Croatian standardology and lexicography**, Milica Mihaljević, Lana Hudeček and Kristian Lewis discuss which COVID-19 neologisms collected from media corpora and online sources should become part of general language dictionaries. They distinguish between Croatian neologisms (single and multiword units) and loanwords and loan translations and stress the importance of responding with prescriptive information in their dictionary to the high number of user questions (regarding orthography, morphology, word formation, usage in a sentence and, last but not least, meaning) concerning all types of neologisms.

Their starting point for the lexicographic description of COVID-19 neologisms was the *Glossary of Coronavirus* compiled by a small group of lexicographers with a clearly descriptive intention: "The purpose of the *Glossary* was to meet the needs of Croatian speakers as soon as possible. It usually records terms as they are used and does not give any normative advice. It includes jargon words as well as scientific

terms which entered the general language". Entries in this glossary were then systematically searched in those corpora that are the basis for the *Croatian Web Dictionary – Mrežnik*, a normative dictionary. The comparison between the differences in the prescriptive vs. the normative approach is informative for other dictionary projects as well.

In Sílvia Barbosa and Susana Duarte Martins' paper **The neologisms of the COVID-19 pandemic in European Portuguese: From media to dictionary**, we learn about the occurrence of COVID-19 neologisms in the press and social networks and whether and how European Portuguese dictionaries have incorporated them. The authors focus on four candidates: *COVID-19*, *coronavirus*, *pandemia*, and *tele-*, and demonstrate with many examples how these are incorporated into new morphological formations, illustrating how vital the lexical neology process in the domain of COVID-19 in a rather short period of time (2020/2021) actually was.

This study also sheds light on how online dictionaries find different ways of reacting to sudden vocabulary expansion, but also on how the Portuguese language was adapted to the new situation by all its speakers. The authors state (what is true also for many of the examples given in other chapters in this volume): "Only the future will tell whether the creative linguistic phenomenon that emerged from the pandemic will persist in the Portuguese language (namely the loss of the neologism status of particular units while being incorporated in the current language lexicon) or whether it will be a source of occasionalisms circumscribed in time and space while the COVID-19 outbreak lasts."

Ieda María Alves, Beatriz Curti-Contessoto, and Lucimara Costa present data on Brazilian Portuguese COVID-19 terminology in their paper **COVID-19 terminology and its dissemination to a non-specialised public in Brazil**. Their corpus-based study "aims to detect, analyse and discuss the characteristics of COVID-19 terminology, in particular the role of the adjective *novo* [*new*] in this terminology, the high recurrence of terms in the plural and the resemantisation of some of the terminological units used".

Their ultimate goal is to create a "terminological dictionary aimed at non-specialised readers in the medical field with little formal education", in which the terms will be presented onomasiologically. As the intended user group comprises a high percentage of functionally illiterate people, the terms will be defined using plain language. The paper exemplifies the manifold lexicographic problems arising when dealing with new terms from the Coronavirus pandemic in such a setting.

In their paper **Neoterm or neologism? A closer look at the determinologisation process,** Rute Costa, Margarida Ramos, Ana Salgado, Sara Carvalho, Bruno Almeida, and Raquel Silva focus on new lexical units in the Portuguese media discourse and their formation, categorization, and lexicographic description. Especially words formed with *covid-* are collected and analyzed regarding the question "whether these words can be considered neoterms or, on the contrary, if having a term in their formation

corresponds to a false neological intuition. In the latter case, rather than a neoterm, we have a neologism resulting from a process of determinologisation."

The authors also discuss several issues regarding the inclusion of such "neoterms" in dictionaries, for example their definition and which domain label should be used. In a template proposal for a lexicographic entry, the authors present their ideas and reflect on the "dictionary as a language model," giving "descriptive guidance" to its users.

In some ways different, but also comparable, problems arise for Sign Language and its lexicographic description, as shown by Mireille Vale and Rachel McKee in their paper **Neologisms in New Zealand Sign Language: A case study of COVID-19 pandemic-related signs**. New signs for suddenly very frequently used new terminology regarding COVID-19 had to be created, conventionalized and disseminated throughout the community of Deaf people in New Zealand. The authors also aim "to explore how and when such neologisms could be entered in the ODNZSL" (*Online Dictionary of New Zealand Sign Language*). The data on signs related to COVID-19 was collected from two sources: signs that were contributed to *NZSL Share*, a web-based platform where users can upload sign videos etc., and signs used by interpreters (e.g. while translating TV briefings on the Corona pandemic). To form the new signs, different strategies were used including "semantic extension; coinage of new words through language-internal mechanisms such as derivation or compounding; and drawing on language-external resources, as calques or direct loans".

Regarding the lexicographic treatment of such lexical innovations, similar problems to conventional language arise, as signs to be included into a dictionary should be fixed, used over a longer period of time outside the original context and widely throughout the whole Deaf community. Using a crowdsourcing platform like *NZSL Share* seems to be a promising tool to find and spread Sign neologisms that then help to update *ODNZSL*.

In the closing paper of this volume, **Using Wiktionary revision history to uncover lexical innovations related to topical events: Application to Covid-19 neologisms**, Franck Sajous explores how data from current revisions in Wiktionary (here demonstrated with the English and the French versions) can be explored to find candidates for COVID-19 neologisms for inclusion in other dictionaries (in addition to exploring media corpus data), thus enabling lexicographers "to monitor, analyse and report quickly a sudden inflow of lexical changes". After explaining his methodology (data processing, ranking new and existing headwords, and annotation of headwords), the author presents his results. Here, readers learn about the different contributor types and how existing and new entries are ranked quarterly and annually, as well as on false negatives.

The study is "based on the hypothesis that Wiktionary's most heavily modified articles can help detect new and existing headwords that are related to topical events", which could be validated for COVID-19 neologisms, at least regarding the English and French Wiktionary version with very active online communities. It

remains to be seen, however, whether the method described here will be able to detect lexical innovations related to topical events with a smaller impact than the Corona pandemic evidently had.

Overall, the findings of the studies in this volume focus on how lexicographic work regarding COVID-19 neologisms has been done and could be improved, either by exploring corpora and other data more systematically, by incorporating users' expertise into the lexicographic process, or by learning from the lexicographic practice of existing dictionaries. Many authors also stress how strongly lexicographic work was affected by the COVID-19 pandemic and its repercussions on the vocabularies of languages around the world in many different ways, but also how, due to such challenges, steps were taken to improve lexicographic work. We hope that the discussion regarding these and other questions related to lexicography and neology in the context of the COVID-19 pandemic and beyond will continue and that this volume contributes to it in a fruitful way.

We would like to express our gratitude to the editorial board of *Lexicographica. Series Maior*, primarily Stefan Schierholz, who quickly accepted our proposal for this volume into the series. And we thank the authors, at the heart of this publication, for their contributions and trust in us.

Bibliography

Klosa-Kückelhaus, Annette/Kernerman, Ilan (2020): Global Viewpoints on Lexicography and Neologisms: An Introduction. In: *Dictionaries. Journal of the Dictionary Society of North America* 41.1: 1–9.

Klosa-Kückelhaus, Annette/Kernerman, Ilan (2021): Converging Lexicography and Neology. In: *International Journal of Lexicography* 34.3: 277–281.

Danica Salazar, Kate Wild

The *Oxford English Dictionary* and the language of Covid-19

1 Introduction

Since the beginning of 2020, the Covid-19 pandemic has dominated public discourse and introduced a wealth of words and expressions to the general vocabulary of English and other world languages. The lexical adaptation necessitated by this global health crisis has been unprecedented in speed and scope, and in response, the *Oxford English Dictionary* (OED) has continually revised its coverage, publishing special updates of Covid-19-related words in 2020 outside of its usual quarterly publication cycle. This article describes how OED lexicographers have analysed language corpora and other text databases to monitor the development of pandemic-related words and provide a linguistic and historical context to their usage.

2 Neologisms of the Covid-19 pandemic

The principal research tool that OED editors use to track the emergence of new words and senses to be considered for addition to the dictionary is Oxford Languages' monitor corpus of English (henceforth the Oxford Monitor Corpus), which currently contains over 14 billion words of web-based news content from 2017 to the present day, and is updated each month. Once a word is identified from the corpus as a candidate for inclusion, editors carefully research both print sources and digital text databases to make sure that there are various independent examples of the word being used, for a reasonable amount of time and reasonable frequency in the types of text in which one would normally expect to find it (see Diamond 2015). There is no exact timespan and frequency threshold for inclusion, as this may vary depending on the type of word. Some words are added to the OED after a relatively short period of time because of their huge social impact, and this has never been truer than in the case of perhaps the most important new word to come out of the Covid-19 pandemic – the word *Covid-19* itself.

Danica Salazar, Oxford University Press, Great Clarendon St, Oxford OX2 6DP, United Kingdom, e-mail: danica.salazar@oup.com
Kate Wild, Oxford University Press, Great Clarendon St, Oxford OX2 6DP, United Kingdom, e-mail: kate.wild@oup.com

https://doi.org/10.1515/9783110798081-002

2.1 The term *Covid-19*

Several of the lexical innovations that emerged during the pandemic are completely new words, or neologisms, the most notable being the name given to the disease at the root of the crisis. *Covid-19* first appears in a situation report published by the World Health Organization (WHO) on 11 February 2020 as the official name of the disease caused by the virus provisionally called *2019 novel coronavirus* (*2019-nCov*) and later formally named *severe acute respiratory syndrome coronavirus 2* (*SARS-CoV-2*). *Covid-19* is shortened from *coronavirus disease 2019*, and follows the WHO's recently adopted best practices for naming new human infectious diseases (WHO 2015). *Covid-19* is an accessible term that indicates the disease's causal pathogen and year of first detection, while avoiding certain geographic, ethnic, cultural, or occupational references that could lead to stigmatizing associations with a particular place or group of people, as had happened in earlier pandemics (e.g., *Spanish flu*, *gay cancer* for AIDS) (Deang/Salazar 2021).

In the months following its coinage, *Covid-19* underwent an exponential rise in usage rarely seen by lexicographers in such a short period of time. By April 2020, when it was added to the OED, it was one of the top five nouns in the dictionary's monitor corpus data for that month (the other four were *people, time, year,* and *coronavirus*), and by May 2020 it had overtaken the word *coronavirus* in frequency (see Figure 1).[1]

One of the challenges with adding extremely new words to dictionaries is that usage may be unfixed. Stewart (2020) explains that when *Covid-19* was first added to the OED in April 2020, it was defined as 'an acute respiratory illness', but in another special dictionary update in July 2020, this definition was changed to 'a disease . . . characterized mainly by fever and cough, . . . capable of progressing to pneumonia, respiratory and renal failure, blood coagulation abnormalities, and death', in order to reflect new information about the effects of the virus on multiple organ systems.

Another aspect of usage which may be subject to change is spelling. There has been quite a lot of discussion online, especially earlier in the pandemic, about whether *Covid-19* should be spelled with an initial capital (as in this article) or with full capitals, *COVID-19*, and different official bodies and news organizations follow different practices. This is the kind of information for which people often turn to a dictionary. Corpus frequencies help to show the most typical use, and in this case it has been found that there is considerable regional variation (see Figure 2): the form with only initial capital is more frequent in the United Kingdom, while there is a clear preference for the all-capital form in the United States. English speakers in Ireland, New Zealand, and South Africa also lean towards the initial-capital form, while those

1 Throughout this article, charts show frequencies per million tokens. (Tokens are the smallest units of a corpus, typically either words or punctuation marks.) Also, variant spellings and inflected forms are included: for example, figures for *Covid-19* include those for *Covid19*, *COVID-19*, etc. (unless stated otherwise, as in Figure 2); figures for *frontliner* include those for *frontliners*, *front-liner*, and so on).

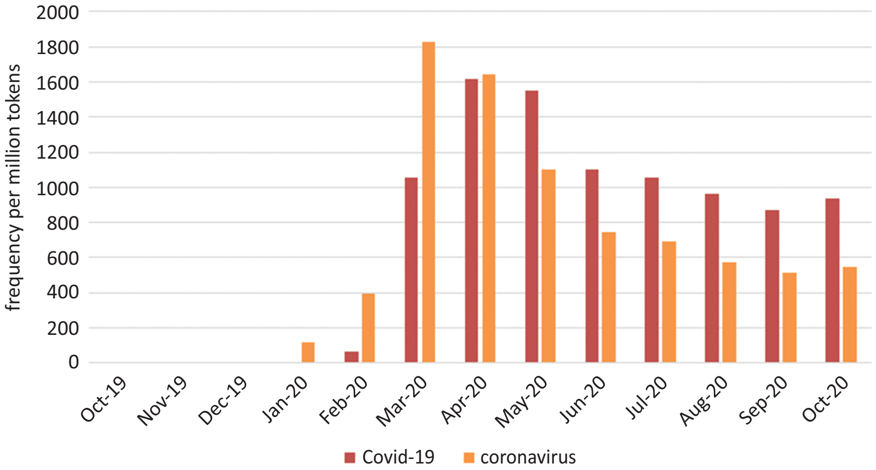


Figure 1: Frequency of *Covid-19* and *coronavirus* in the Oxford Monitor Corpus, October 2019 to October 2020.

in Canada, Australia, and India prefer the all-capital form. There may be fluctuations as time goes on, and this is something the OED will continue to track. Following its usual style, the OED entry gives the most common British form as the headword but lists the other forms as variants.

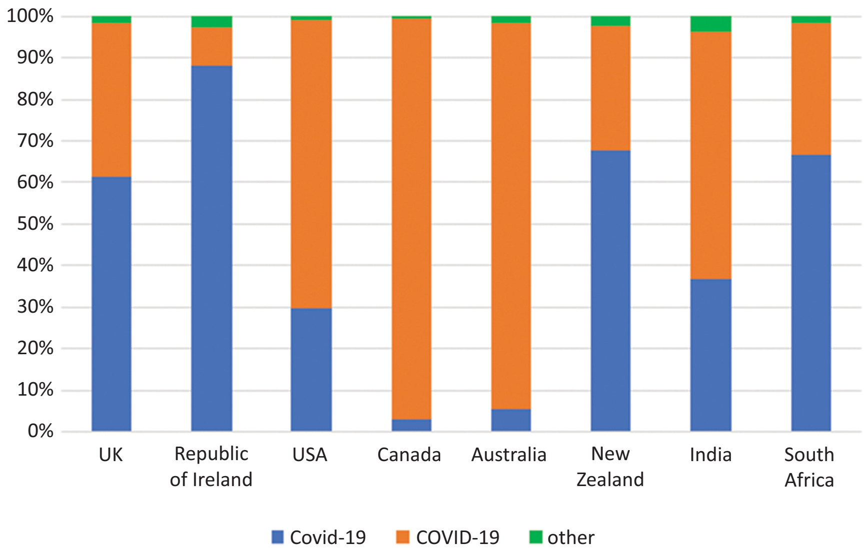


Figure 2: Relative frequencies of *Covid-19*/*COVID-19* in selected varieties of English in the Oxford Monitor Corpus, as of July 2021. “Other” includes *CoVID-19*, *Covid19*, etc.

2.2 Other neologisms

In addition to *Covid-19* itself, numerous other new words have entered the English language as a result of the pandemic. These include shortenings denoting Covid-19 or the coronavirus, like *Covid* (first recorded in a tweet on the same day that *Covid-19* was coined), *corona*, *rona* (particularly frequent in colloquial US and Australian use), *C-19*, and *nCoV*.

While pandemic-related words such as *lockdown* have been widely borrowed from English into other languages (see section 6), borrowing into English seems to have been a less common source of new vocabulary during the time of Covid-19. There are exceptions, such as *Hamsterkauf*, a German word meaning 'panic buying' (from the idea of a hamster – German *Hamster* – hoarding food in its cheeks) and occasionally used as an English word early in 2020: examples from the Oxford Monitor Corpus include "supermarkets are experiencing a wave of hamsterkauf" and "the initial hamsterkauf phase of the pandemic". However, most of the uses refer to the word as a loanword from German, and *hamsterkauf* seems not to have taken root as a naturalized English word.

Much more productive methods of neology during the pandemic have been blending and compounding. *Covid-* and *corona-* have been particularly productive elements, especially in *covidiot* and also in more ephemeral formations like *covidivorces* 'divorces prompted by the stress of lockdown' or, more positively, *coronials* 'the generation of babies born during lockdown' (from *Covid* and *millennials*). As face-to-face interactions began to be prohibited and the videoconferencing software *Zoom* became ubiquitous, coinages such as *Zoombombing*, *Zoom-ready*, and *zumping* 'dumping someone over Zoom' emerged. (The OED also added an entry for the use of *Zoom* as a verb.) There have also been various coinages formed with *-demic* (from *pandemic* or *epidemic*), such as *twindemic*, referring to a hypothetical pair of pandemics occurring at the same time, and sceptical formations like *scamdemic* and *plandemic*. Many other new blends and compounds – some serious, some more playful – have been created, including *anthropause* (the global slowdown of travel and other human activity during the pandemic), *pancession* (an economic recession caused by a pandemic), *isodesk* (a home workplace), *maskne* (acne caused by wearing a face mask), and a plethora of words denoting alcoholic drinks consumed during lockdown or self-isolation, such as *quarantini* and *locktail*.

Although some of these words have experienced widespread popularity, many are likely to be quite short-lived, and have not yet been added to the OED, but the dictionary's editors will continue to track their development. Criteria for determining whether and when to add a word to the OED include longevity, frequency of occurrence, and variety of sources, although there are no rigid rules and each word is considered on its own merits (see further Diamond 2015).

3 Words with new senses or new significance

Not all of the lexical developments during the pandemic have been completely new words – in fact, our corpus monitoring has shown that most of them are existing words that have developed new senses or gained special significance as a result of the pandemic.

3.1 Corpus keywords

Table 1 shows the top ten keywords in the Oxford Monitor Corpus for the first seven months of 2020. Keywords are words that appear significantly more frequently in one part of the corpus than in the corpus as a whole (Kilgarriff 2009), so these are words that were particularly frequent in the given months. Keywords relating to the coronavirus crisis are highlighted in bold in the table, where it can be observed that *Covid-19*, *Covid*, and other abbreviations are the only actual neologisms. The other keywords are words with a longer history, like *coronavirus*, *lockdown*, *pandemic*, *furlough*, and *covering*. A list such as this is used by OED lexicographers to check against the dictionary's coverage so as to determine whether any new information needs to be added to existing entries.

Table 1: Top 10 keywords in the Oxford Monitor Corpus, January to July 2020.

Jan 2020	Feb 2020	Mar 2020	Apr 2020	May 2020	June 2020	July 2020
bushfire	**Covid-19**	**Covid-19**	**PPE**	**reopen**	defund	**covering**
coronavirus	**coronavirus**	**pandemic**	**lockdown**	**lockdown**	Juneteenth	**Covid**
Iranian	**quarantine**	**distancing**	**pandemic**	**Covid-19**	brutality	**in-person**
SARS	**pandemic**	**coronavirus**	**ventilator**	**pandemic**	anti-racism	**mask**
Iraqi	**virus**	**self-isolate**	**stay-at-home**	**Covid**	racism	**mask-wearing**
sign-stealing	**outbreak**	**lockdown**	**Covid-19**	**distancing**	**Covid**	**pandemic**
koala	caucus	**self-isolation**	**furlough**	**hydroxy-chloroquine**	Confederate	**distanced**
virus	locust	**sanitiser/ sanitizer**	**distancing**	**covering**	looting	**Covid-19**
impeachment	**infect**	**quarantine**	**coronavirus**	**furlough**	**covering**	**SARS-CoV-2**
airstrike	**epicentre/ epicenter**	**ventilator**	**N95**	**stay-at-home**	kneel	**pre-pandemic**

The list of keywords presents a fascinating overview of changing global events and concerns in the first seven months of 2020. In January and February, some of the keywords were related to the coronavirus; others referred to different world events such as the Australian bushfires, the assassination of Qasem Soleimani, Donald Trump's impeachment, the Democratic caucuses, locust swarms in East Africa, and investigations into the Astros sign-stealing scandal. In March, however, every one of the top ten keywords was in some way related to coronavirus. This remained the case until June, when many of the keywords reflected the impact of the Black Lives Matter movement and the protests following the killing of George Floyd on 25 May 2020.

It is also revealing to compare the pandemic-related keywords in the table. In January 2020, the words mainly related to naming and describing the virus: *coronavirus, SARS, virus*. By March, April, and May the keywords reflected the social and economic impact of the virus: *social distancing*, *self-isolation*, *quarantine*, and *lockdown* were all especially frequent, as was *furlough*, following the introduction of the UK's Coronavirus Job Retention Scheme in late March. Issues surrounding the medical response are reflected in the keywords *PPE* (for *personal protective equipment*) and *ventilator*.

In May 2020, we saw the first signs of looking ahead to life post-lockdown, with *reopen* as the top keyword. This trend continued in July, when there was an interesting pattern of contrast with virtual life as people started thinking about or tentatively restarted face-to-face interaction: *in-person* increased in frequency, used in contexts which previously would not normally have been necessary (since the "in-person" version was the norm), as in *in-person worship* and *in-person graduation*.

In July 2020, the top keyword was *covering*, overwhelmingly in *face covering* or in other uses referring to face masks (including *facial covering* or simply *covering* in this sense, in examples such as "shop staff do not have to wear coverings"). *Mask* and *mask-wearing* also appeared as keywords in July, reflecting ongoing discussions about when and where masks and coverings should be worn.[2]

Analysis of these keywords prompted a number of additions and updates to the OED. For example, new entries for *self-isolation* (and related terms), *face covering*, and *PPE* were added (although these terms are not completely new – see section 3.3). Some entries were revised to account for shifts in use or emphasis: for example, the relevant sense of *lockdown* was updated to include the public health measure aspect, while the entry for *furlough* was fully revised, including a new comment about the spread of the sense 'dismissal or suspension from employment' from the US to the UK and other countries.

2 The foregoing discussion draws on the analysis in Wild (2020a) and Wild (2020b). Table 1 shows keywords only up to July 2020: after this month, the corpus keyword lists were less dominated by Covid-19-related topics, and reflected other events such as the US presidential elections. However, Covid-19 certainly continued to be a theme, with keywords including *pre-Covid*, and, from the end of 2020 and beginning of 2021, *vaccine* and related words.

3.2 The term *coronavirus*

As noted above, many of the words used in the context of the pandemic are not completely new but were relatively uncommon before 2020. This is the case with the word *coronavirus* itself, the name of the group of enveloped, single-stranded RNA viruses of which the causal pathogen of Covid-19 is a member. The word *coronavirus* is first recorded in the OED in 1968, in an article in *Nature*, but before the Covid-19 pandemic its use was mainly confined to medical and scientific specialists. This is reflected in corpus data: as shown in Figure 1 (section 2.1), *coronavirus* was relatively rare in general news media before 2020; by March 2020, it was dominating the global conversation.

One way of illustrating the extent to which the word *coronavirus* became overwhelmingly frequent at the beginning of the pandemic is to compare its frequency with that of other significant words at the time. Figure 3 compares the frequency of *coronavirus* with that of words referring to other major news topics in 2019 and 2020 – *climate*, *Brexit*, and *impeachment* – and shows that *coronavirus* was over ten times as frequent as any of these words at its peak. Figure 4 shows that, by March 2020, *coronavirus* was as frequent as one of the most commonly used nouns in the English language, *time*.

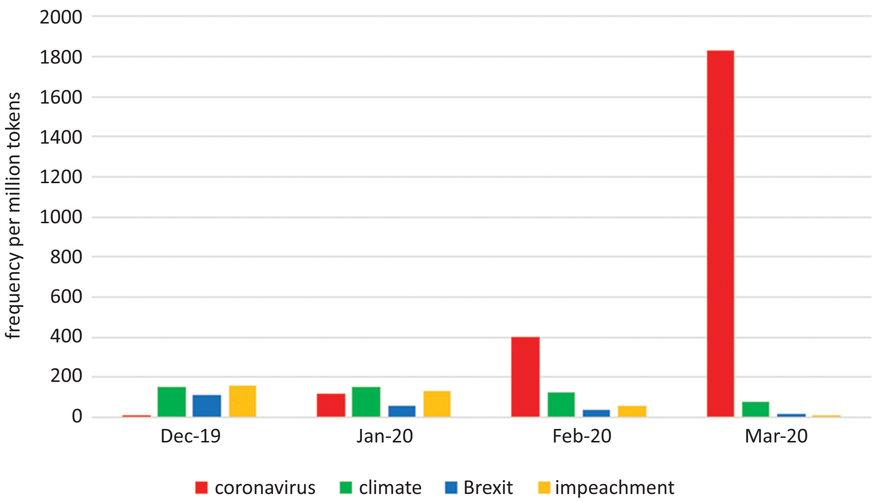


Figure 3: Frequency of *coronavirus, climate, Brexit,* and *impeachment* in the Oxford Monitor Corpus, December 2019 to March 2020.

As with *Covid-19*, changes have been made to the OED's entry for *coronavirus* in light of developments during the pandemic (Stewart 2020). A second sense has been added to refer specifically to those coronaviruses that cause life-threatening diseases in humans, including SARS (Severe Acute Respiratory Syndrome), MERS

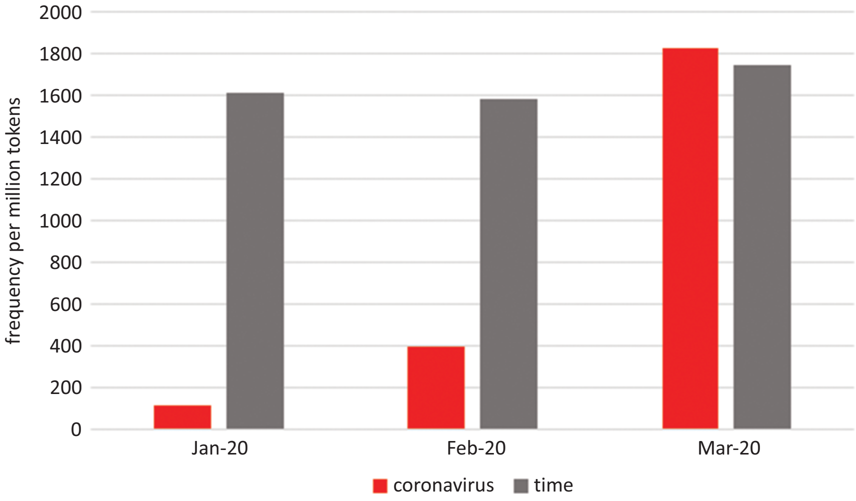


Figure 4: Frequency of *coronavirus* and *time* in the Oxford Monitor Corpus, January 2020 to March 2020.

(Middle East Respiratory Syndrome), and Covid-19. Additionally, since the name of a disease often also ends up being applied to the pathogen causing it, and vice versa, both *Covid-19* and *coronavirus* began to be used interchangeably for the disease and the virus; again, this has been reflected in the updated *OED* entries.

3.3 Older words newly added to the *OED*

Social distancing was one of the entries added to the OED in its first special pandemic update in April 2020. As can be seen from Figure 5, this term saw an enormous increase in usage: its frequency was negligible before 2020; then by April 2020 it was occurring over 250 times for every million tokens in the Oxford Monitor Corpus, which is roughly the same frequency as that of the word *food*.

However, when OED lexicographers researched this word, consulting databases of books, newspapers, journals, and other types of written sources, they found that *social distancing* is far from being a new term. It dates back to 1957, originally signifying an aloofness or deliberate attempt to distance oneself from others socially. It is only decades later when it acquired the now more familiar sense of limiting physical contact in order to avoid infection, but even this sense goes back almost two decades, to 2004. This antedating of what may originally seem to be obvious neologisms is something that often occurs when a lexical item is researched for the OED, and there are several examples of such terms from the Covid-19 pandemic: *self-quarantine* dates back to 1876 as a noun and 1918 as a verb, *elbow bump* to 1902, and *contact tracing* to

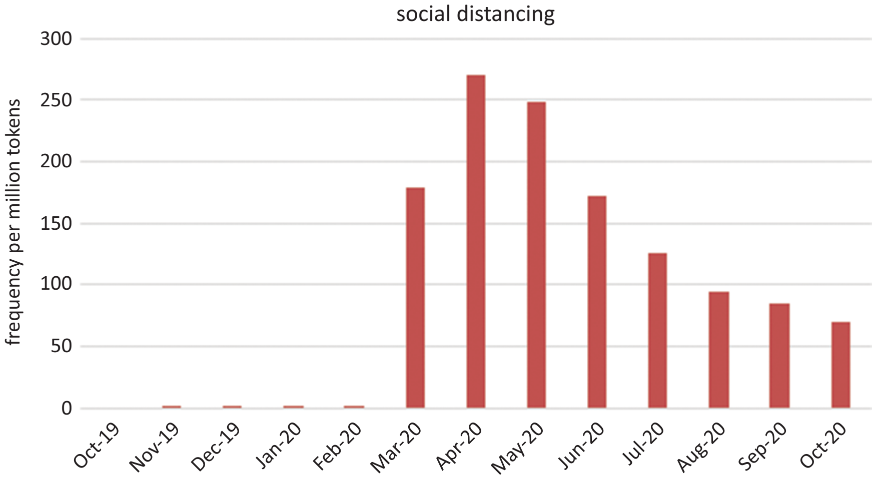


Figure 5: Frequency of *social distancing* in the Oxford Monitor Corpus, October 2019 to October 2020.

1910; *face covering* is first recorded in 1732, and first used in a medical context in 1946. Although these terms were new to the OED when they were added as a result of the pandemic, they are not completely new to the language.

Some expressions were coined during previous public health crises or for other kinds of emergencies but have been revived during the time of Covid-19. *Infodemic*, a blend of *information* and *epidemic*, was coined in 2003 during the SARS epidemic to refer to the outpouring of often unsubstantiated information relating to a crisis, and was then also widely used to describe the proliferation of news around coronavirus. The phrase *shelter-in-place*, a protocol instructing people to find a place of safety in the location they are occupying until the all clear is sounded, was devised as an instruction for the public in 1976 in the event of a nuclear or terrorist attack but was then adapted as advice to people to stay indoors to protect themselves and others from disease (Paton 2020).

3.4 New senses or nuances of existing words

Collocational information gleaned from corpus data helps OED lexicographers understand the contexts surrounding the usage of a word and discover particular nuances or senses. For example, in the dictionary's entry for *frontline*, the sense of the adjective as used in *frontline worker/employee/staff*, etc., had been defined as 'of a person: working at the forefront of an organization's public activity, typically as the point of direct contact with customers, clients, users of the organization's services, etc.' This was an accurate summary when the entry was first revised a few years ago, but the

focus of the sense has shifted during the Covid-19 pandemic. OED editors compared salient collocates of *frontline* in 2020 with those of previous years and found that although some had remained unchanged – *frontline staff* is one consistently common collocation – others, such as the following, stood out as much more frequent in 2020: *frontline* ***nurse/medic/caregiver***; *frontline* ***healthcare/health-care workers***; *frontline* ***warrior/hero***; ***courageous/heroic*** *frontline* ***workers***; ***essential*** *frontline* ***worker***. This very positive sentiment associated with frontline workers, and the focus on such workers as carrying out essential roles, especially in health care, led to the OED definition being expanded as follows: 'of a person: working at the forefront of an organization's public activity, typically as the point of direct contact with customers, clients, users of the organization's services, etc., (now) *esp*. designating such an employee who provides a service regarded as vital within the community, such as a health-care worker, teacher, etc.; often in *frontline worker*'.

Another new pandemic-related sense added to an existing OED entry is *bubble*. *Bubble* is a word with a long history, its literal sense dating to the Middle Ages and various figurative senses (mainly relating to either impermanence or protection) to the Early Modern period. In 2021 a new sense was added, 'a group consisting of a restricted number of people who have a close relationship or regular social contact; (later) *spec*. such a group whose members are, under public health measures, permitted to be in close physical proximity'. The first, general strand of this sense dates back to 2000, but the OED definition notes that the later specific strand 'arose in 2020 as part of the official recommendations of some governments in response to the Covid-19 pandemic'. Again, the emerging new sense is discernible from corpus data: a comparison of collocates in 2020/2021 and previous years highlights new or newly significant collocations such as ***household*** *bubble* and ***support*** *bubble*.

4 The spread of scientific terminology in general discourse

Another notable feature of the language of the pandemic has been the way that it has introduced scientific terms into general discourse. As both scientists and the public endeavoured to increase their understanding of the coronavirus and its effects, specialist scientific and medical language became increasingly prominent. This development has already been discussed with reference to the term *coronavirus* (see section 3.2), and it is reflected in numerous other words in various fields.

Lexical items from the field of epidemiology that were previously known mainly to the scientific and medical community were suddenly being heard in the news and in everyday conversation. For example, *reproduction number*, *reproductive number*, *R number*, or simply *R*, became widely used as people became preoccupied with "getting the R down". This crossing over from specialist to general vocabulary

is reflected in the set of quotations in the OED entry for this sense of *R*: the earliest use is from the proceedings of an epidemiology conference published in 1975, while the most recent example is from a news article.

Other epidemiological terms such as *community transmission*, *community spread*, *case fatality rate*, and *flattening the curve* became widespread enough to merit inclusion in the OED's first special pandemic update published in April 2020. In its second special update three months later, the dictionary focused even more on scientific and medical terminology, adding such terms as *cytokine storm* 'an overactive immune response occurring in various infectious and non-infectious diseases, characterized by the excessive production of cytokines and resulting in intense localized or generalized inflammation'; *spike protein* 'a glycoprotein projecting from the envelope which binds to a receptor on the host cell and facilitates entry of the viral genome into the host cell', and *CPAP* 'continuous positive airway pressure, a method of respiratory therapy in which air at a pressure higher than atmospheric pressure is pumped into the lungs through the nose or nose and mouth during spontaneous breathing', a less invasive treatment for Covid-19 patients than one involving a ventilator (Stewart 2020). Again, these terms are not completely new, but they have become widely familiar to non-specialists as a result of the pandemic.

5 Regional variation

We have discussed the use of corpus data to identify new words, spikes in frequency, and shifts in collocation and other aspects of usage. Corpora also provide useful information about the distribution of a word in different varieties of English, which is reflected in the labelling and metadata of new or revised dictionary entries.

In the case of *self-isolate*, *self-quarantine*, and related words, OED editors working on these terms felt that although there are technical differences between them, they are often used interchangeably, the main difference being in regional distribution. To confirm this, they looked at various corpora. The clearest picture can be seen in the Coronavirus Corpus, a corpus of news articles relating to Covid-19 (Davies 2019–), which shows that *self-quarantine* is more common in the United States than in Canada, Great Britain, Ireland, Australia, and New Zealand, where *self-isolate* and *self-isolation* are preferred (see Wild 2020b). A note to this effect has been added to the OED's updated entry for *self-quarantine*, v.: 'In recent use, in the context of the Covid-19 pandemic, *self-isolate* and *self-quarantine* have often been used interchangeably, with *self-quarantine* being more common in the United States'.

Corpus frequency data also enabled OED editors to analyse the regional distribution of the word *frontliner*. They discovered that although it is used worldwide, it is particularly frequent in Southeast Asia, especially in the Philippines and Malaysia (see Figure 6); in other countries the more usual term is *frontline worker* or similar.

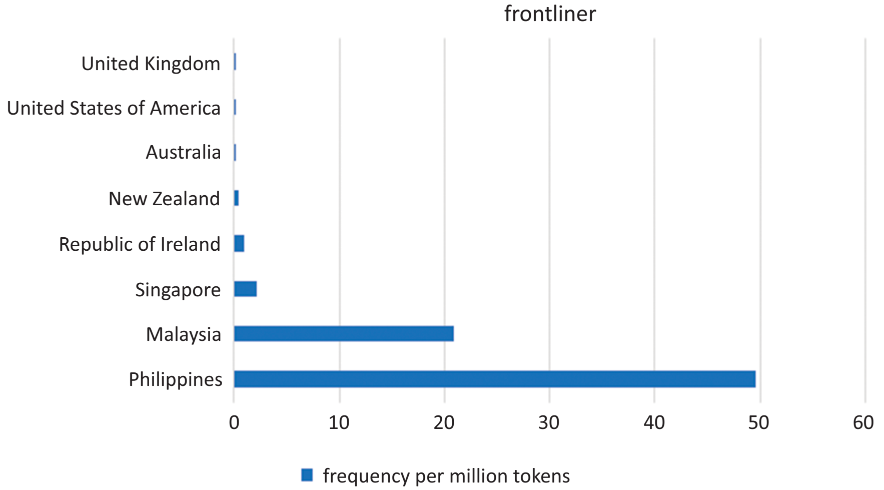


Figure 6: Frequency of *frontliner* in selected varieties of English in the Oxford Monitor Corpus, July 2020.

For this reason, the OED entry for the relevant sense of *frontliner* is labelled "now chiefly *South-East Asian*".

Also showing some interesting geographic variation are the names for the set of measures that many countries have taken to contain the spread of the virus by severely limiting the movement of people outside the home. *Lockdown* is the word with the most widespread use and is the preferred term in countries such as the United Kingdom, Canada, and Australia. In the United States the coronavirus restrictions are called *shelter-in-place*. The word *iso,* short for *isolation*, is also used colloquially, especially in Australia and the United States. In Malaysia, the initialism *MCO* is used, short for *movement control order,* while in the Philippines, *ECQ* is preferred, short for *enhanced community quarantine* – both phrases are the official government designations for these countries' stay-at-home regulations. In Singapore, there was a remarkable spike in usage of the term *circuit breaker* in April 2020 when it was adopted by the Singaporean government as the name for its strict quarantine measures (see Figure 7). Known to most people as a safety device that stops the flow of current in an electric circuit, *circuit breaker* is also familiar to those in finance as a regulatory instrument designed to prevent panic selling by temporarily stopping trading on an exchange. While it makes sense for a global business hub such as Singapore to have adapted a piece of finance slang in such a way, it is noteworthy that later in 2020, in September and October, *circuit breaker* also became a much-used term in British English, describing a short, fixed-term set of restrictions which scientists recommended the government should implement in order to stem another incoming tide of coronavirus infections (see Figure 8).

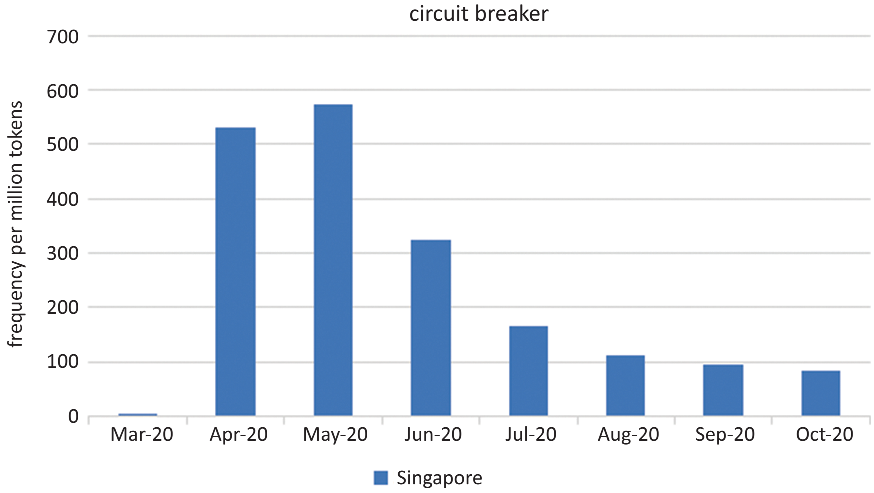


Figure 7: Frequency of *circuit breaker* in the Singapore component of the Oxford Monitor Corpus, March to October 2020.

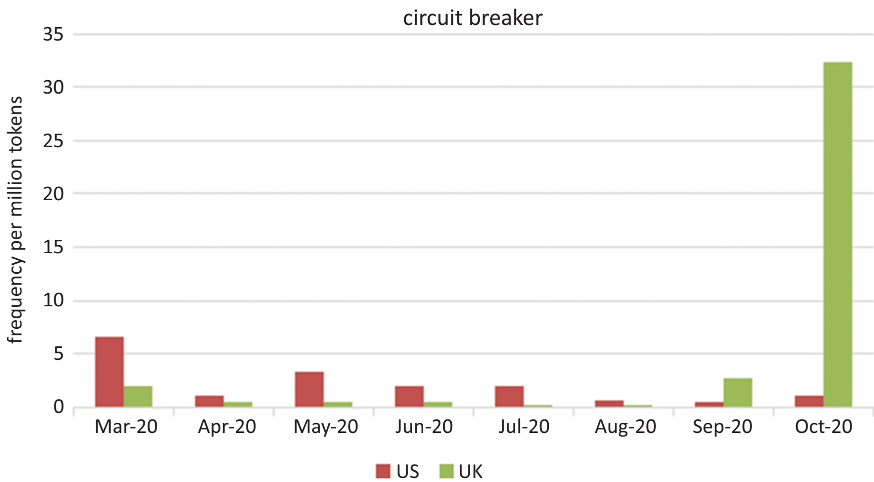


Figure 8: Frequency of *circuit breaker* in the UK and US components of the Oxford Monitor Corpus, March to October 2020.

Local responses to the coronavirus pandemic have also resulted in several neologisms in different varieties of English. In the Philippines, Filipinos from other regions stranded in a locked-down Manila are referred to as *LSI*s, short for *locally stranded individuals*; in Singapore, a person who needs to self-isolate is issued an *SHN* or *stay-home notice*; while in India those who wish to cross internal borders need to have an *e-pass*, an official government document authorizing a person's movement during

quarantine. Australians try to keep themselves safe from the virus through the regular use of *sanny* (hand sanitizer), while West Africans wash their hands using the *Veronica bucket* – a type of sanitation equipment composed of a covered bucket with a tap fixed at the bottom and a bowl fitted below it to collect wastewater, named after its Ghanaian inventor, Veronica Bekoe (Salazar 2020).

A Southeast Asian term added to the OED in 2016 suddenly gained global notoriety at the outset of the Covid-19 pandemic: *wet market*. This term, first attested in 1978, was originally used only in Southeast Asian countries to refer to a market for the sale of fresh meat, fish, and produce, an essential part of the region's food supply chain. However, the identification of a Wuhan market as ground zero for the coronavirus outbreak led people outside of Southeast Asia to incorrectly conflate wet markets with illegal wildlife markets, subjecting wet markets to much public criticism (Lim 2020) and causing a considerable increase in the usage of the term in the early months of 2020 (see Figure 9).

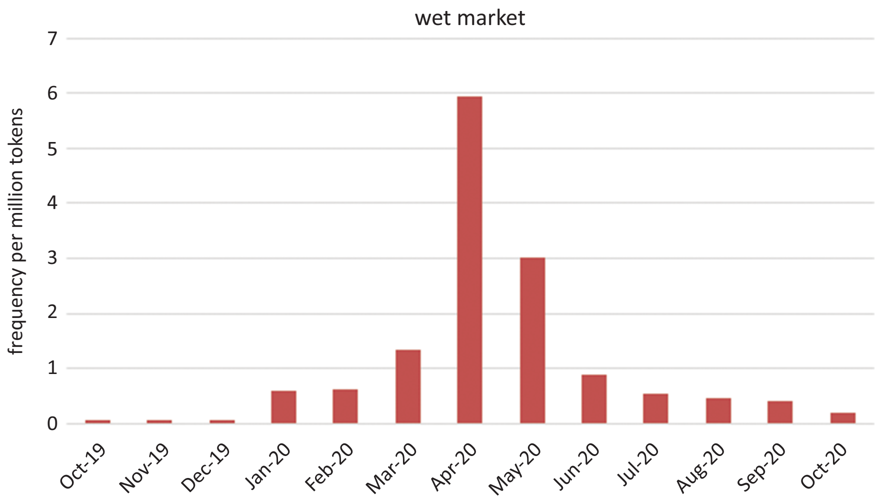


Figure 9: Frequency of *wet market* in the Oxford Monitor Corpus, October 2019 to October 2020.

6 Languages other than English

The lexical monitoring carried out by Oxford Languages' lexicographers has informed the coverage of the pandemic lexicon not only in the OED, but also in other Oxford dictionaries of current English and even in Oxford dictionaries of other languages. Key terms related to Covid-19, such as the neologisms and newly prominent words mentioned throughout this article, were translated into 19 different languages by Oxford University Press editors and translators in Oxford and in its international offices

in China, India, East Africa, and South Africa so that new words and senses could be incorporated into its monolingual and bilingual dictionaries of these languages. These translations into Afrikaans, Arabic, Catalan, Chinese, Dutch, Filipino (Tagalog), French, German, Hindi, Italian, Northern Sotho, Portuguese, Setswana, Spanish, Swahili, Tamil, Telugu, Xhosa, and Zulu have also been made freely available as downloadable resources online.[3]

The translation of key coronavirus words into such diverse languages has also provided important insights into the impact of the pandemic on these languages. There are some commonalities with English: the emergence of new words and senses, the increased significance of medical and scientific terminology, and the prominence of expressions referring to government and individual actions aimed at containing the spread of the virus and mitigating its social and economic effects. There are also interesting differences. For example, the English word *lockdown* has been borrowed by several languages including Dutch, Filipino, German, Italian, and Telugu, while other languages prefer their equivalent forms for *confinement*, for example, *confinament* for Catalan, *confinement* for French, *confinamento* for Portuguese, and *confinamiento* for Spanish. Some languages use corresponding expressions conveying closure, for instance, إغلاق *'iighlaq* for Arabic, 封锁 *fēngsuǒ* and 封闭 *fēngbì* for Chinese and *ukuvalwa thaqa kwezwe* for Zulu.

The Covid-19 translation project has also highlighted the influence of English, the principal language of global scientific communication, on the Covid-19 vocabulary of these languages. This influence can be seen in some notable lexical innovations. In Italian, for instance, the word *droplet* has come to refer not only to the very small airborne drops of secretions from the nose, throat, or lungs by which the coronavirus can be transmitted, but also to the distance one person must maintain from another to prevent such a transmission from happening.

7 Conclusion

The OED's efforts to document the lexical change brought to the English language by the coronavirus pandemic continued throughout 2020, culminating with the *Words of an Unprecedented Year* report, which was published at the end of the year in place of the usual selection of a single Word of the Year (Oxford Languages 2020). This expansive report on the words that defined 2020 features an entire section dedicated to the language of Covid-19.

However, the work did not end there. Further pandemic-related additions and revisions to the dictionary have been included in the OED's regular quarterly

---

3 The translations can be downloaded from Oxford Languages' Covid-19 Language Hub: https://languages.oup.com/covid-19-language-resources/#translations (last access: 12 August 2021).

updates in 2021, with *face shield*, *essential worker*, *mask up*, and the aforementioned *bubble* being notable examples. Several more are scheduled to be published in upcoming updates. OED lexicographers will continue to monitor their in-house corpora and other language data to identify and document new words and senses associated with the pandemic that have had such an impact on our language and our lives.

Bibliography

Davies, Mark (2019–): The Coronavirus Corpus. [https://www.english-corpora.org/corona/; last access: 13 July 2020].

Deang, Richard Karl/Salazar, Danica (2021): *Major health crises and the OED: language evolution and challenges in health communication*. Oxford. [https://public.oed.com/blog/major-health-crises-and-the-oed/; last access: 11 August 2021].

Diamond, Graeme (2015): Making decisions about inclusion and exclusion. In: Durkin, Philip (ed.): *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press, 532–545.

Kilgarriff, Adam (2009): Simple maths for keywords. In: Mahlberg, Michaela/González-Díaz, Victorina/Smith, Catherine (eds.): *Proceedings of the Corpus Linguistics Conference*. University of Liverpool (no page numbering).

Lim, Lisa (2020): Why Asia's wet markets are being unfairly targeted in bad pandemic press. In: *South China Morning Post Magazine*, 25 April. [https://www.scmp.com/magazines/post-magazine/short-reads/article/3081363/why-asias-wet-markets-are-being-unfairly/; last access: 12 August 2021].

OED = *Oxford English Dictionary* (2000–). Third edition (in progress). Ed. John A. Simpson (2000–2013) and Michael Proffitt (2013–). Oxford: Oxford University Press. [https://www.oed.com; last access 15 August 2021].

Oxford Languages (2020): *Words of an unprecedented year*. Oxford. [https://pages.oup.com/ol/word-of-the-year-2020; last access: 12 August 2021].

Paton, Bernadette (2020): *Social change and linguistic change: the language of Covid-19*. Oxford. [https://public.oed.com/blog/the-language-of-covid-19/; last access: 12 August 2021].

Salazar, Danica (2020): *Circuit breakers, PPEs, and Veronica buckets: World Englishes and Covid-19*. Oxford. [https://public.oed.com/blog/circuit-breakers-ppes-veronica-buckets-world-englishes-covid-19/; last access: 12 August 2021].

Stewart, Trish (2020): *July 2020 update: scientific terminology of Covid-19*. Oxford. [https://public.oed.com/blog/july-2020-update-scientific-terminology-of-covid-19/; last access: 11 August 2021].

Wild, Kate (2020a): *Corpus analysis of the language of Covid-19*. Oxford. [https://public.oed.com/blog/corpus-analysis-of-the-language-of-covid-19/; last access 12 August 2021].

Wild, Kate (2020b): *Using corpora to track the language of Covid-19: update 2*. Oxford. [https://public.oed.com/blog/using-corpora-to-track-the-language-of-covid-19-update-2/; last access 12 August 2021].

World Health Organization (2015): *Best practices for the naming of new human infectious diseases*. [https://apps.who.int/iris/bitstream/handle/10665/163636/WHO_HSE_FOS_15.1_eng.pdf; last access: 11 August 2021].

Annette Klosa-Kückelhaus

German Corona-related neologisms and their lexicographic representation

1 Introduction

Between January 2020 and July 2021, many new words and phrases contributed to the expansion of the German vocabulary to enable communication under the new conditions that evolved during the Covid-19 pandemic. Medical and epidemiological vocabulary was integrated into the general language to a large extent. Suddenly, some lexemes from general language were used with very high frequency, while other words were used less often than before. These processes of language change can be studied in various ways, for example, in corpus linguistics with respect to the frequency or emergence of certain words in certain types of texts (e.g. press releases vs. posts in social media), in critical discourse analysis with respect to certain participants of the discourse (e.g. vocabulary of Covid-19 pandemic deniers),[1] or in conversation analysis (e.g. with respect to new verbal interactions in greetings and farewells). The rapid expansion of vocabulary has notably affected also lexicography as a discipline of applied linguistics.

General language dictionaries or terminological dictionaries have quickly reflected on how the German lexicon evolved during the Covid-19 pandemic. For example, new entries have been added to *Digitales Wörterbuch der Deutschen Sprache*, a comprehensive synchronic general language dictionary of German, such as *Corona-party*[2] 'a privately organized party held during the corona crisis, bypassing the coronavirus containment measures'. New senses were also recorded, such as the 'highest school-leaving certificate based exclusively on school achievements already achieved [. . .] without taking final examinations' attributed to the noun *Durchschnittsabitur* which has the older meaning 'university entrance qualification with a mediocre, average final grade point average'. Duden publishing house with its extensive monolingual online dictionary on contemporary German, *Duden online*, added new entries like *Covid-19*[3] focusing on spelling variants (*Covid-19* in general and mostly *COVID-19* in technical language) and grammatical information (e.g. in the entry *Coronavirus*:[4]

1 See Wengeler/Roth (2020) for several studies on the corona discourse and also Weinert (2021).
2 https://www.dwds.de/wb/Coronaparty (last access: 10 June 2022).
3 https://www.duden.de/suchen/dudenonline/Covid-19 (last access: 10 June 2022).
4 https://www.duden.de/suchen/dudenonline/Coronavirus (last access: 10 June 2022).

Annette Klosa-Kückelhaus, Leibniz Institute for the German Language, Mannheim, Germany, e-mail: klosa@ids-mannheim.de

https://doi.org/10.1515/9783110798081-003

in terminology mostly neutral gender: *das Coronavirus*, in everyday conversation outside of terminology also masculine gender: *der Coronavirus*). Regarding technical lexical items, for example, the German *Bundessprachenamt* ('Federal Language Bureau') published an online dictionary, *Coronavirus Terminology in German, English, French, Dutch, Polish, Russian, Spanish*, comprising 18,490 terms (e.g. *akute respiratorische Erkrankung* 'acute respiratory infection', *Bewegungsprofil* 'movement profile').

The following sections, however, will focus on the ways in which a German neologism dictionary project has chosen to capture and document lexicographic information in a timely manner. Both challenges and advantages arise from lexicographic practice "at the pulse of time". The *Neologismenwörterbuch* is presented as an example that lends itself well to such a discussion because its subject (neologisms) is characterized as new, innovative, and constantly changing.

2 German neologisms and neologism dictionaries

New words emerge in the German language all the time, but not all of them are neologisms. Being a morphologically productive language, many new compounds or derivations are used only once; and such nonce words are not lexicalized. Neologisms, on the other hand, are defined here as lexical units or senses/meanings that emerge in a language community over a specific period of time of language development, which (a) diffuse, (b) are generally accepted as language norms, and (c) are perceived by the majority of speakers as new for some time (cf. Herberg et al. 2004: XII). There are some indicators for the acceptance of a neologism in German: its increasing overall frequency, its distribution in many different text types/genres, and its use in many different discourses. Other indicators tell us how far the process of lexicalization of new words has developed (cf. Lemnitzer 2010: 69):

- Pragmatic criteria: A neologism is no longer written in quotation marks, hedge words (e.g. *sogenannt* 'so-called') are not used any more, and distancing phrases (e.g. *wie man heute sagt* 'as we say today') are abandoned.
- Grammatical criteria: The gender of nouns is invariable; a full conjugational paradigm for verbs has developed.
- Word-formation criteria: A neologism is used as first and second component in an increasing number of compound nouns; a borrowed neologism is combined with indigenous lexemes in word-formation products.

According to the definition given above, some time (possibly a couple of years) must pass before a new word can be classified as a neologism, and it will only continue to be part of this word class for some time (possibly a couple of years) before it is no longer perceived as new. This notion directly impacts the lexicographic description of these lexemes in neologism dictionaries, which are defined as specialized reference

guides focusing on the description of meaning and usage of such lexemes in a specific language which became part of the vocabulary of that language at a certain time (cf. Barnhart/Barnhart 1990; Lemnitzer 2010; Wiegand 1990 for more details). There are some (neologism) dictionaries that only record neologisms retrospectively, that is after their lexicalization. The *Neologismenwörterbuch* is a typical, corpus-based example of this type of dictionary, Quasthoff (2007) is another. Other neologism dictionaries, like *Die Wortwarte*, record neologisms at their "moment of birth" (cf. Lemnitzer 2010: 67) before they are fully lexicalized, but are nevertheless accepted parts of the lexicon.

Regarding the remarkably quick extension of the German lexicon during the Covid-19 pandemic, that is, in a relatively short time span of less than two years, the second type of neologism dictionaries can continue expanding its list of entries according to established criteria of lemma inclusion. The other, retrospective type of neologism dictionary, however, cannot react directly to the expansion of the lexicon, as lexicographers do not know yet whether the new lexemes, phrases or senses/meanings will become generally established parts of the lexicon after some time.

3 User needs and lexicographic responses

Many people around the world noticed new words emerging in their languages in the context of the Covid-19 pandemic, including internationalisms like *corona*, *Covid-19*, *social distancing* as well as other language-specific ones. Journalists, teachers, staff of medical or political institutions, etc., reacted to the general interest in these new lexical items and soon started to publish glossaries with definitions and/or some encyclopedic information. As early as March 2020,[5] several daily newspapers (e.g. *Süddeutsche Zeitung*, *Die Rheinpfalz*), news magazines (e.g. *Der Spiegel*), radio stations (Bayerischer Rundfunk, Deutschlandfunk), and news programmes (Tagesschau) in Germany offered corona glossaries with information on neologisms, medical terminology, etc., to their audience (cf. Möhrs 2020). Many of these are still available online (June 2022), and some have been updated since then. In addition, several scientific organizations and other public institutions began to publish (online) terminological resources (e.g. *Corona-Glossar* by the Helmholtz Association of Research Centres, or the above-mentioned multilingual terminological Corona database by

5 In Germany, the first attested infection with the SARS-CoV-2 virus was registered on 27 January 2020, and by the middle of March 2020, infections were registered in all federal states. General information on the Covid-19 pandemic in Germany can be found in *Bundeszentrale für politische Bildung* (2021).

Bundessprachenamt). Finally, some individuals utilized social media to call for collaboration regarding the collection of corona words.[6]

The first German dictionary project which quickly reacted to the challenge of offering information on those words from the medical and epidemiological contexts that suddenly were of general interest was *Digitales Wörterbuch der Deutschen Sprache* (DWDS). By the middle of March 2020, the *DWDS-Themenglossar zur COVID-19-Pandemie* ('Thematic glossary on the covid-19 pandemic') was published,[7] and it has since been updated continuously.[8] Duden publishing house updated their online dictionary on contemporary German extensively in August 2020, and then more Corona neologisms were added regularly.[9]

In the following sections, more information is given on how the *Neologismenwörterbuch* ('Neologisms Dictionary') project at IDS Mannheim reacted to the unprecedented expansion of the German lexicon with new lexemes, phrases and meanings related to the pandemic.

4 The *Neologismenwörterbuch*

4.1 General remarks

The *Neologismenwörterbuch* covers new words and new senses or meanings established since 1991. New entries are compiled continuously, and the reference guide is published online as part of the dictionary portal OWID (Online-Wortschatz-Informationssystem Deutsch, 'Online information system on the German lexicon') of the Leibniz Institute for the German Language at Mannheim.[10] In this dictionary project, the editorial interpretation (evaluation of print and online media) is combined with a quantitative corpus-linguistic method[11] to extract candidates for

---

6 See, for example, the threads of tweets by Nadja Hahn, https://twitter.com/nadjasnews/status/1334517401359015936 (December 2020), or Lara Fritzsche, https://twitter.com/larafritzsche/status/1304330059935895552 (September 2020) (last access: 10 June 2022).

7 https://www.dwds.de/themenglossar/Corona (last access: 10 June 2022).

8 At the same time, DWDS had included new corona-related entries to its comprehensive general language dictionary of contemporary German and had updated several entries regarding the corona pandemic. For examples, see above.

9 According to a private communication by Kathrin Kunkel-Razum, editorial director of Duden dictionaries, in October 2021.

10 For the decades of 1991–2000 and 2001–2010, print dictionaries are also available (Herberg/Kinne/Steffens 2004; Steffens/al-Wadi 2015). The lexicographic concept for these volumes goes back to the late 1980s (cf. Heller/Herberg/Lange/Schnerrer/Steffens 1988; Kinne 1989) and 1990s (cf. Herberg 1997 and 1998).

11 For further information on the editorial and corpus-linguistic methods applied to detect potential neologisms for the *Neologismenwörterbuch* cf. Klosa/Lüngen (2018).

inclusion into the dictionary (semi-)automatically. The *Neologismenwörterbuch* contains entries on:

- single words, e.g., *Spam* ('spam'),
- multi-word expressions, e.g., *Generation Facebook* ('age group of young people for whom the Internet and, in particular, communication via social networks are a matter of course'),
- new word formation elements e.g., *[. . .]holic* (used to form nouns denoting someone who is very intensely involved with something, resembling an addiction to something).
- new senses of existing words in German, e.g., *texten* ('to text').

Since its first online publication in 2005 and for the next 15 years, new and comprehensive entries were regularly uploaded to the dictionary by the end of each year. Full entries in the *Neologismenwörterbuch* comprise information on etymology, orthography, pronunciation, meaning, usage, grammar, word formation, encyclopedic information, illustrations, as well as frequency and emergence in the corpus. In 2020, a new type of a more concise entry structure was established for all the words that do not need more extensive information because, for example, no pragmatic restrictions apply to their use. As such, they offer details on grammar, orthography, meaning, word formation or etymology as well as the decade of emergence. These entries are uploaded in thematically related groups on a monthly basis (e.g. April 2021: agriculture, July 2021: crime, September 2021: fashion).[12] The latest inclusions, covering terms used in agriculture, crime or fashion, illustrate how new words often center around a specific new subject. By providing such thematically related bundles of new items users get a compact overview of recent lexical-semantic developments within a specific area.

The *Neologismenwörterbuch* offers different lists of lemma groups[13] as well as a thematic search[14] and other extended search options[15] for both lemma types. For example, for any neologism dating back to a specific decade (1991–2000, 2001–2010, or 2011–2020), all new phrases, every new word formation element, or all neologisms assigned to the domain of sports, medicine, or fashion, etc., can easily be found.

12 For an overview see https://www.owid.de/docs/neo/listen/kurzartikel.jsp (last access: 10 June 2022).
13 See https://www.owid.de/docs/neo/wortartikel.jsp for different lemma groups (last access: 10 June 2022).
14 See https://www.owid.de/docs/neo/gruppen.jsp?grp=1 for thematic search option (last access: 10 June 2022).
15 See https://www.owid.de/docs/neo/suche/index.jsp for extended search options (last access: 10 June 2022).

While dictionary users thus already had many options to access extensive information on neologisms that are accepted parts of the German lexicon in *Neologismenwörterbuch*, the dictionary – according to its definition of a neologism (see above) – however, did not offer any information regarding new words that are not fully lexicalized yet. Therefore, users of the dictionary were often not able to obtain details on words that were particularly conspicuous at a particular time in a specific discourse, thus raising questions concerning their meaning, correct spelling, etc. This, however, did not imply that the lexicographers had not already collected these words with some preliminary information in a list of candidates for inclusion in an internal database. Consequently, the *Neologismenwörterbuch* project started to publish a list of monitored words (*Wörter unter Beobachtung* 'Monitored Words'[16]) in March 2020. All entries in this list consist of lexical units that emerged since 2011, for which only time will tell whether they will diffuse and manifest as language norms. For each of these words, only a (preliminary, rough) explanation of meaning is given, and usage is illustrated by one or two corpus citations, wherever relevant, there are hyperlinks to more encyclopedic information (e.g. in Wikipedia) and the date of recording the word is noted. When items from this inventory are described in either comprehensive or concise entries, they are removed from the list, which is updated quarterly.

By the middle of April 2020, the first corona-related neologisms were included in this online collection of monitored words to react quickly to the sudden demand for information on meaning and usage of new words in the context of the pandemic. Around 30 entries on words such as *Contacttracing* ('contact tracing'), *Coronababy* ('child conceived during exit and contact restrictions in Covid-19 pandemic (in home quarantine)' and 'child of a Covid-19 patient'), or *zoomen* ('to communicate and work with Zoom® video conferencing software') were published. Soon it became obvious that the number of Corona-related neologisms would exceed the number of other monitored words by far, so a separate list on the Corona-related vocabulary[17] was published at the end of April 2020 (with a little over 60 entries). Since then, this list (see Figure 1 with the entry on *No-Covid-Strategie*, 'no Covid strategy', i.e. 'concept aimed at slowing down or containing the Covid-19 pandemic as completely as possible through appropriate measures (e.g. pushing the infection figures to zero and thus creating virus-free zones)') was updated every fortnight between April and June 2020 and from then on, on a monthly basis. In October 2021, the list included more than 1,800 Corona-related neologisms, and still, more than 700 candidates in an internal database awaited lexicographic description and inclusion into the online index.

---

16 See https://www.owid.de/docs/neo/listen/monitor.jsp (last access: 10 June 2022).

17 See https://www.owid.de/docs/neo/listen/corona.jsp# (last access: 10 June 2022).

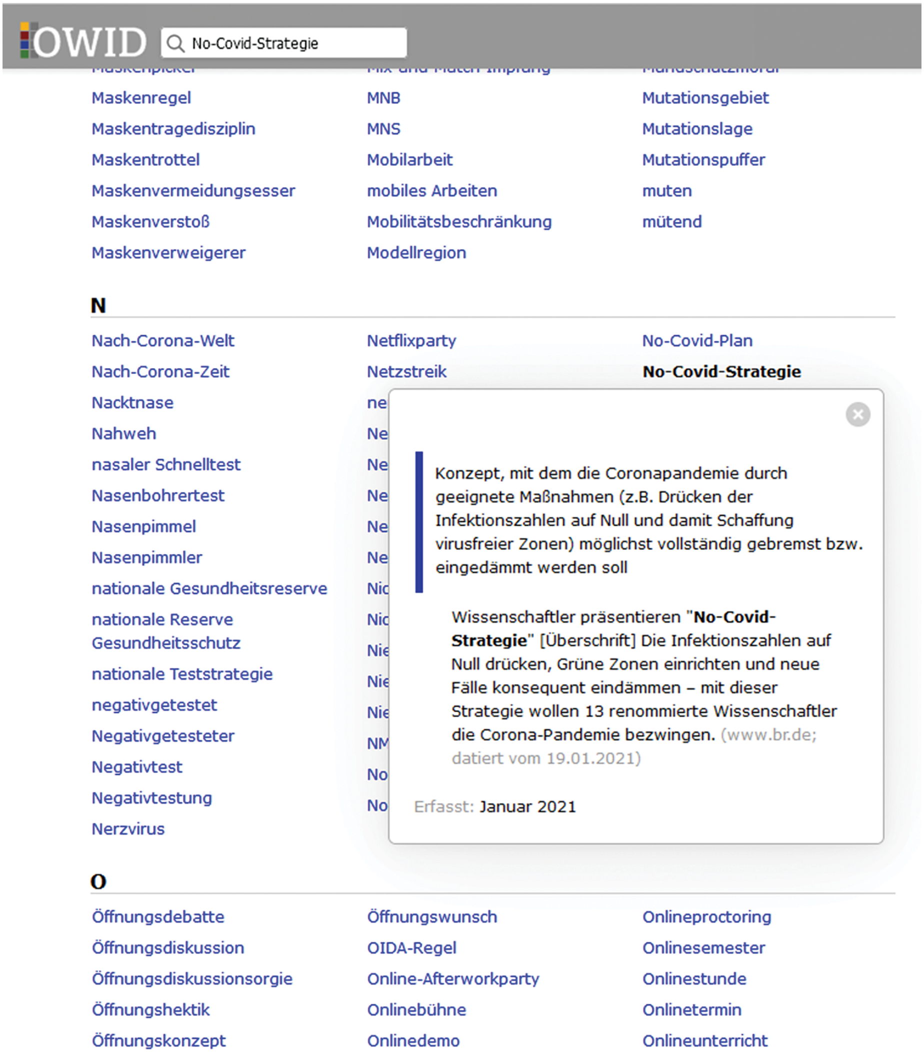


Figure 1: Extract from the dictionary index "Neuer Wortschatz rund um die Coronapandemie" ('New vocabulary around the Covid-19 pandemic') in the *Neologismenwörterbuch*.

Generally, (semi-)automatic retrieval of neologism candidates (see above) in the Deutsches Referenzkorpus – DeReKo ('German Reference Corpus') of IDS[18] was used to detect candidates for the list of new vocabulary around the Covid-19 pandemic. In

18 See https://www.ids-mannheim.de/digspra/kl/projekte/korpora/ (last access: 10 June 2022).

addition, the IDS corpus tool cOWIDplus Viewer[19] and its newer version OWIDplusLIVE[20] were deployed, giving access to RSS-feeds of thirteen German online newspapers and magazines in weekly updates. The editorial collection of candidates through press reading and, for example, browsing the daily Twitter trends, was continued as well and several other glossaries and lists of Corona-related words (see above) were evaluated systematically. Finally, many dictionary users participated by sending their suggestions for new words to be included via an online form in the *Neologismenwörterbuch*.[21]

Dictionary users were informed on the expansion of the list via the IDS newsletter *IDS aktuell*[22] (published quarterly). Progress on the compilation of the index is also recognizable by the date of recording given at the bottom of each entry (cf. Figure 1: Erfasst: Januar 2021, 'Added: January 2021'). Furthermore, users learn about the status of the list as work-in-progress by a very short introductory text on the website. For the first time in its history, the team working on *Neologismenwörterbuch* also started to publish a number of shorter lexicological studies on the IDS web page[23] concerning words and specific word groups with special significance within the corona discourse (e.g. *Social Distancing*, *Maske* 'face mask', *Lockdown* and *Shutdown*), in which readers are referred to the index of Corona-related vocabulary where appropriate. In the following section, some examples from this list are discussed to illustrate the lexicographic challenges from expanding the IDS neologism dictionary "at the pulse of time".

4.2 News words

In Germany, as soon as the first vaccine for SARS-CoV-2 was found and officially approved, the vaccination campaign started (in January 2021), with specific groups of people being prioritized over others for medical reasons. Some people were dissatisfied with this solution and experienced a feeling of resentment, called *Impfneid* ('vaccine envy'), which was recorded in *Neologismenwörterbuch* as a typically coined neologism following the German compounding rules (verb *impf[en]* 'to vaccinate' + noun *[der] Neid* 'envy') in January 2021.[24] This feeling then led some people to try to

19 See https://www.owid.de/plus/cowidplusviewer2020 (last access: 10 June 2022)/.
20 See https://www.owid.de/plus/live-2021/ (last access: 10 June 2022).
21 See https://www.owid.de/wb/neo/mail.html (last access: 10 June 2022).
22 See https://www.ids-mannheim.de/aktuell/presse/newsletter/ (last access: 10 June 2022).
23 See https://www.ids-mannheim.de/sprache-in-der-coronakrise/ (last access: 10 June 2022).
24 See https://www.owid.de/docs/neo/listen/corona.jsp#impfneid (last access: 10 June 2022).

be vaccinated before their turn. These people are referred to as *Impfdrängler*[25] or *Impfvordrängler*[26] ('vaccination tailgaters'), for which, yet again, typical new compounds were recorded in the dictionary in January 2021. Still, as Figure 2 illustrates, both words were in use only a short period of time (January to May/June 2021), as by July 2021 enough vaccines were available nationwide and prioritizing immunization was no longer necessary. Thus, both words seem to be short-lived lexemes which may probably not meet the criteria for neologisms to be included in *Neologismenwörterbuch* in the end (see above), but which are (a) justifiably part of a list of words still monitored, and (b) might well be candidates for a specialized dictionary on the corona discourse, as the topic of vaccination is lexically very rich in German and characterized by quite different discourse participants (e.g. opponents and proponents of vaccination, experts and laypeople).

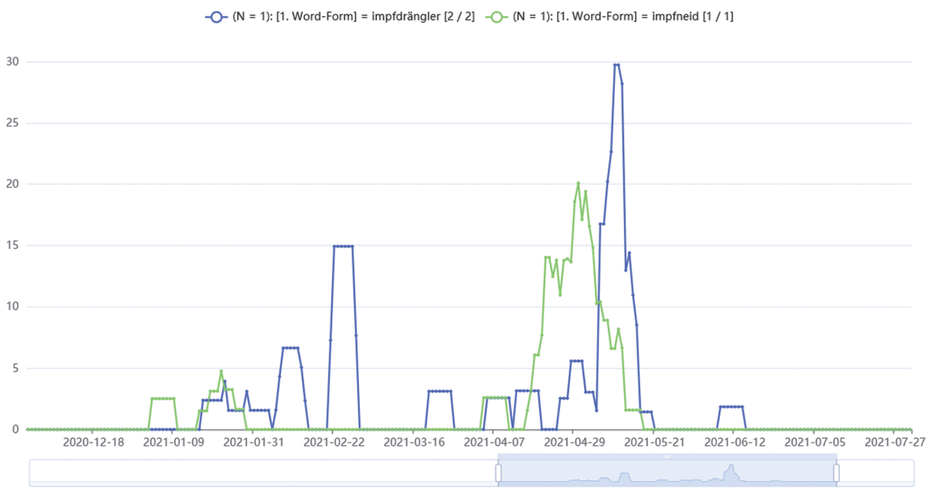


Figure 2: Relative frequency of *Impfdrängler* ('vaccination tailgater') and *Impfneid* ('vaccination envy') in RSS-feeds of 13 German online newspapers and magazines between December 2020 and July 2021.[27]

4.3 New senses

New senses for long-existing lexemes can be attributed to the corona discourse less often than new words and phrases. One example is the German abbreviation *3G*, originally denoting the third generation of mobile telecommunication networks

25 See https://www.owid.de/docs/neo/listen/corona.jsp#impfdraengler (last access: 10 June 2022).
26 See https://www.owid.de/docs/neo/listen/corona.jsp#impfvordraengler (last access: 10 June 2022).
27 Source: "OWiDplusLIVE", see https://www.owid.de/plus/live-2021/ (last access: 10 June 2022).

(also called *UMTS*). In the context of the Covid-19 pandemic, however, this abbreviation refers to two different notions:

a) three circumstances that should be avoided during the Covid-19 pandemic to avoid an increase of infections: ***g**eschlossene Räume* ('enclosed spaces'), ***G**edränge* ('crowding') and ***G**espräche ohne Abstand* ('conversations without distance'),
b) three groups of people with special health status who pose a low risk of infection in terms of contracting the SARS-CoV-2 virus: ***G**etestete* ('tested persons'), ***G**eimpfte* ('vaccinated persons'), ***G**enesene* ('persons recovered from Covid-19').

Both senses are illustrated in the entry on *3G*[28] in *Neologismenwörterbuch* (recorded in June 2021). Since September 2021, the expression *2G*[29] is used to refer to only two groups, namely vaccinated or recovered persons. By now, both abbreviations are already part of numerous German compounds (cf. Figure 3). Only the future will tell whether the new senses of *2G* and *3G* will possibly be used with a wider reference to methods of fighting an epidemic caused by any virus.

"Stay home"-Sticker	200er-Inzidenz	3G-Modell
"Stopp Corona"-App	200er-Notbremse	3G-Nachweis
"Wir bleiben zuhause"-Sticker	2G	3G-Regel
1-2-3-Regel	2G-Modell	4-Tage-R
15-km-Beschränkung	2G-Option	50-Neuinfektionen-Marke
15-km-Grenze	2G-Optionsmodell	501Y.V2
15-km-Radius	2G-Regel	50er-Inzidenz
15-km-Regel	2G-Regelung	50er-Marke
15-km-Regelung	30-Sekunden-Regel	7-Tage-R
15-Minuten-Regel	3G	

Figure 3: Extract from the list of neologisms around the Covid-19 pandemic in *Neologismenwörterbuch* showing compounds with *2G* and *3G*.[30]

4.4 New uses

Another interesting example is the word (*der*/*das*) *Coronavirus* (noun, masculine or neutral gender), which has been attested in German since the 1980s in the context of AIDS research. Since then, the use of the word shows strong conspicuities in that

28 See https://www.owid.de/docs/neo/listen/corona.jsp#3g (last access: 10 June 2022).
29 See https://www.owid.de/docs/neo/listen/corona.jsp#2g (last access: 10 June 2022).
30 See https://www.owid.de/docs/neo/listen/corona.jsp# (last access: 10 June 2022).

it does not continuously occur with approximately the same frequency or with a continuously increasing or decreasing tendency in the texts of the “German Reference Corpus – DeReKo”, but it shows two prominent frequency peaks, namely in the years 2003 and 2013 (cf. Figure 4). The significantly more frequent occurrence of *Coronavirus* in 2003 is due to the SARS infection wave discussed at that time, and the high rate in 2013 is due to quite a few known MERS cases at that time. Of course, a next, much more significant peak will show for the years 2020 and 2021. This word is thus a good example of how current events of the day affect the frequency with which words are used. For *Neologismenwörterbuch*, however, the term *Coronavirus* itself is not a suitable candidate for inclusion as, on the one hand, it had been attested long before the beginning of the Covid-19 pandemic and, on the other hand, the dictionary does not cover any periods earlier than 1991 (see above).

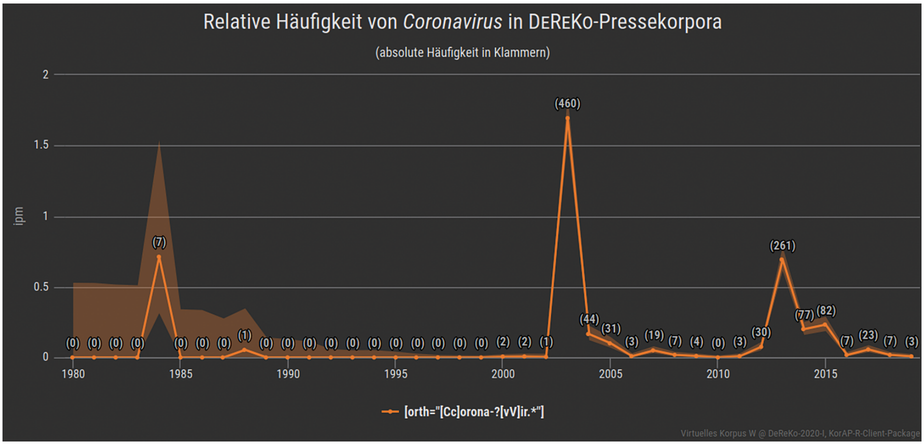


Figure 4: Relative frequency of *Coronavirus* in the “German Reference Corpus – DeReKo” between 1980 and 2019 (with absolute frequency numbers in brackets).[31]

In contrast, the noun *Corona* as an abbreviation of *Coronavirus* is documented in *Neologismenwörterbuch* with a short entry[32] with three different senses (‘the SARS-CoV-2 virus’, ‘the disease caused by SARS-CoV-2, i.e. Covid-19’, ‘the Covid-19 pandemic and the crisis caused hereby’) and other lexicographic information. In this case, the editorial team argued that there is no need to monitor the word further because of its frequency and communicative relevance.

31 Graph by Mark Kupietz, project “Ausbau und Pflege der Korpora geschriebener Gegenwartssprache” at IDS Mannheim, see https://www.ids-mannheim.de/digspra/kl/projekte/korpora/ (last access: 10 June 2022).
32 See https://www.owid.de/artikel/408108 (last access: 10 June 2022).

Finally, the entry of the noun *(das) Homeoffice* with its two senses 'workspace at home' and 'working from home' illustrates how new uses originating in the Covid-19 pandemic led to the revision of existing entries in *Neologismenwörterbuch*. *Homeoffice* in both senses is widely used in German since the mid-1990s, but the aspect of using modern telecommunication channels to do so was added to the definitions in July 2020 and new phrases like *im Homeoffice arbeiten* ('to work from home') or the synonym *Heimbüro* 'home office' now supplement the entry due to their increased frequency in the Covid-19 pandemic.

Over all, the list of new words and expressions around the Covid-19 pandemic in *Neologismenwörterbuch* contains predominantly single word entries (e.g. (*das*) *Autokonzert*[33] 'live performance of an artist, music group, or the like, where the audience listens to the music over a special radio channel while sitting in their car') and not more than 10% are multi-word expressions (e.g. (*die*) *Generation Corona*,[34] 'age group of young people particularly affected by the economic consequences of the Covid-19 pandemic, who have poor prospects of a successful start to their working lives due to the crisis'). More than 90% of the new lexemes are nouns (e.g. (*der*) *Maskomat*[35] 'vending machine (in front of a store) where individually packaged masks can be purchased'). Some adjectives (e.g. *postpandemisch*[36] 'concerning the period after the Covid-19 pandemic') and (very few) verbs (e.g. *teamsen*[37] 'to communicate, work, hold classes, etc., over the Internet using the Teams® video conferencing software') make up for the rest.

Most of the German corona-related vocabulary is formed following German word formation rules, i.e.:

– compounding, e.g. *Doppelmutante*[38] ('(presumably) highly contagious variant of SARS-CoV-2 virus with two genome alterations'), from *doppelt* 'double' + *(die) Mutante* 'mutant',
– derivation, e.g. *downlocken*[39] ('to restrict economic and social activities (to contain an epidemic)'), derived from (*der*) *Lockdown* + suffix *-en*,
– abbreviation, e.g. *Delta*[40] ('(presumably) highly contagious mutation of the SARS-CoV-2 virus, which was detected for the first time in India'), from *(die) Deltavariante*.

---

33 See https://www.owid.de/docs/neo/listen/corona.jsp#autokonzert (last access: 10 June 2022).
34 See https://www.owid.de/docs/neo/listen/corona.jsp#generation-corona (last access: 10 June 2022).
35 See https://www.owid.de/docs/neo/listen/corona.jsp#maskomat (last access: 10 June 2022).
36 See https://www.owid.de/docs/neo/listen/corona.jsp#postpandemisch (last access: 10 June 2022).
37 See https://www.owid.de/docs/neo/listen/corona.jsp#teamsen (last access: 10 June 2022).
38 See https://www.owid.de/docs/neo/listen/corona.jsp#doppelmutante (last access: 10 June 2022).
39 See https://www.owid.de/docs/neo/listen/corona.jsp#downlocken (last access: 10 June 2022).
40 See https://www.owid.de/docs/neo/listen/corona.jsp#delta (last access: 10 June 2022).

Only 5–10 % of the words on the list are English loanwords (e.g. (*das*) *Weaning*,[41] 'slow weaning of an intensive care patient from mechanical ventilation'). Compounding is the predominant word coining process (over 90%), while derivation or abbreviation are used only in a few cases (for examples see above).

While the relation between single word and multi-word entries in *Neologismenwörterbuch* as a whole is roughly the same (90%: 10%) and the proportions of parts of speech are similar as well (nouns: 83%, verbs: 8%, adjectives: 3 %, other: 6 %), the *Neologismenwörterbuch* as such contains approximately 61% of words that are formed in German, while 39% are loanwords. Those neologisms that are formed in German are compounds (70%), derivatives (21%), or abbreviations (9%). One reason for the deviation of the numbers for the corona-related neologisms in the list from the data of *Neologismenwörterbuch* as a whole is that the corona-related neologisms are still monitored. Thus, whole synonyms clusters are part of the inventory, for example:

– *Coronateststation*,[42] *Coronateststelle*,[43] *Coronatestzentrum*,[44] *COVID-19-Teststation*,[45] *COVID-19-Teststelle*,[46] *COVID-19-Testzentrum*,[47] all denoting a mobile walk-in or drive-in facility for people to be tested for a possible infection with the SARS-CoV-2 virus, or
– *Immunitätsausweis*,[48] *Immunitätsnachweis*,[49] *COVID-19-Ausweis*,[50] *Covidpass*,[51] *digitaler Impfpass*,[52] *grüner Pass*,[53] all designating an official document confirming the immunity of a person with respect to the SARS-CoV-19 virus.

As of October 2021, it remains unclear which lexeme of each group will eventually be the most dominant term and thus the most eligible candidate to be entered into the reference guide, while the less frequent lexemes are only treated inclusively under the corresponding qualified headword. For the lexicographic catalogue of monitored words of *Neologismenwörterbuch*, the general rule applies that semantically transparent compounds or derivatives are explained in the dictionary, rule of

41 See https://www.owid.de/docs/neo/listen/corona.jsp#weaning (last access: 10 June 2022).
42 See https://www.owid.de/docs/neo/listen/corona.jsp#coronateststation (last access: 10 June 2022).
43 See https://www.owid.de/docs/neo/listen/corona.jsp#coronateststelle (last access: 10 June 2022).
44 See https://www.owid.de/docs/neo/listen/corona.jsp#coronatestzentrum (last access: 10 June 2022).
45 See https://www.owid.de/docs/neo/listen/corona.jsp#covid-19-teststation (last access: 10 June 2022).
46 See https://www.owid.de/docs/neo/listen/corona.jsp#covid-19-teststelle (last access: 10 June 2022).
47 See https://www.owid.de/docs/neo/listen/corona.jsp#covid-19-testzentrum (last access: 10 June 2022).
48 See https://www.owid.de/docs/neo/listen/corona.jsp#immunitaetsausweis (last access: 10 June 2022).
49 See https://www.owid.de/docs/neo/listen/corona.jsp#immunitaetsnachweis (last access: 10 June 2022).
50 See https://www.owid.de/docs/neo/listen/corona.jsp#covid-19-ausweis (last access: 10 June 2022).
51 See https://www.owid.de/docs/neo/listen/corona.jsp#covidpass (last access: 10 June 2022).
52 See https://www.owid.de/docs/neo/listen/corona.jsp#digitaler-impfpass (last access: 10 June 2022).
53 See https://www.owid.de/docs/neo/listen/corona.jsp#gruener-pass (last access: 10 June 2022).

thumb that is also less strictly applied for comprehensive and concise entries. The deviations in the data will be studied more closely in the future, when the development of usage and its manifestation can be examined in the IDS corpus data in more detail.

5 Conclusion

Neologism dictionaries seem to be predestined to react quickly to recent and possibly rapid developments of the lexicon of a given language, as it is their objective to cover new lexemes and phrases, a phenomenon naturally associated with change. However, this can only be the case if the dictionary in question aims at documenting new words close to their moment of creation, taking into account the risk that some of the neologisms recorded may only be short-lived. In contrast, retrospective neologism lexicography runs the risk of offering details about new words at a time where the needs of users for this information have decreased and the words in question may no longer be considered as novel. In a situation like the Covid-19 pandemic, where the lexicon of languages around the world (including German) has expanded to a large extent in a very short amount of time and the usage frequency of some words has changed perceptibly in contrast to former times, this question has accumulated even greater importance. In this study, the solution found for *Neologismenwörterbuch* of IDS was demonstrated, i.e., by offering both types of information in one online reference work. On the one hand, fully described neologisms in comprehensive entries are being compiled for words that emerged in German in a specific period of time of language development (the 1990s, the 2000s, the 2010s, the 2020s) and which have diffused and are now generally accepted; and on the other hand, some headwords are still monitored and are being collected, receiving only brief and essential semantic specifications.

Work in the *Neologismenwörterbuch* project exceeded lexicographic practice by additionally offering corpus-based, evidence-based research results on neologisms around the Covid-19 pandemic to the public. The editorial team regularly published short essays on specific word groups (between April 2020 and August 2021) and got into contact with dictionary users by email correspondences based on an online word proposal form as part of the dictionary. The editors also gave many interviews to newspapers, magazines, radio, and TV stations[54] that followed the developments and the emergence of new coinings around the Covid-19 pandemic with great interest. Besides regular continuous editorial work, all these activities meant additional

54 See https://www.ids-mannheim.de/aktuell/presse/online-presse/ for the press review of IDS, last access: 10 June 2022.

workloads for a rather small team of lexicographers,[55] but which helped raising the general awareness for the lexicographical work at IDS and for the usefulness of reliable, up-to-date dictionaries, making it worthwhile to immerse in lexicography "at the pulse of time".

Bibliography

Dictionaries

Corona-Glossar (2020–). Ed. by Helmholtz Association of Research Centres [https://www.helmholtz.de/glossar/corona-glossar/, last access: 10 June 2022].

Coronavirus Terminology in German, English, *French*, Dutch, *Polish*, Russian, *Spanish*. Ed. by Bundessprachenamt [https://app.coreon.com/5ea2adb797e1040100eb7ff3, last access: 10 June 2022].

Die Wortwarte – Reloaded. Wörter von heute und morgen (2000–). Ed. by Lemnitzer, Lothar [https://wortwarte.org/, last access: 10 June 2022].

Duden online (no year). Ed. by Dudenredaktion [https://www.duden.de/woerterbuch, last access: 10 June 2022].

DWDS = *Digitales Wörterbuch der Deutschen Sprache* (2003–). Ed. by Berlin-Brandenburg Academy of Sciences and Humanities [https://www.dwds.de, last access: 10 June 2022].

Herberg, Dieter/Kinne,Michael/ Steffens,Doris (2004): *Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen*, in collaboration with Tellenbach, Elke and al-Wadi, Doris. Berlin/New York: de Gruyter.

Neologismenwörterbuch (2005–). Ed. by Leibniz Institute for the German Language [https://www.owid.de/docs/neo/start.jsp, last access: 10 June 2022].

OWID = *Online-Wortschatz-Informationssystem Deutsch* (2008–). Ed. by Leibniz Institute for the German Language [https://www.owid.de/, last access: 10 June 2022].

Quasthoff, Uwe (2007): *Deutsches Neologismenwörterbuch. Neue Wörter und Wortbedeutungen in der Gegenwartssprache*. Berlin/New York: de Gruyter.

Steffens, Doris/al-Wadi, Doris (2015): *Neuer Wortschatz. Neologismen im Deutschen 2001–2010*, 3rd ed. Mannheim: Institut für Deutsche Sprache.

References

Barnhart, Robert K./Barnhart, Clarence L. (1990): The Dictionary of Neologisms. In: Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (eds.): *Wörterbücher – Dictionaries – Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Berlin/New York: de Gruyter, 1159–1166.

Bundeszentrale für politische Bildung (2021): *Corona. Pandemie und Krise*. Aus Politik und Zeitgeschichte, 10714, Bonn: Bundeszentrale für politische Bildung.

55 The editorial team of *Neologismenwörterbuch* in 2020 and 2021 consisted of approximately 1.5 full-time equivalents.

Heller, Klaus, et al. (1988): *Theoretische und praktische Probleme der Neologismenlexikographie. Überlegungen und Materialien zu einem Wörterbuch der in der Allgemeinsprache der DDR gebräuchlichen Neologismen*. Berlin: Zentralinstitut für Sprachwissenschaft.
Herberg, Dieter (1997): Neologismen im allgemeinen Wörterbuch oder Neologismenwörterbuch? Zur Lexikographie von Neologismen. In: Konerding, Klaus-Peter/Lehr, Andrea (eds.): *Linguistische Theorie und lexikographische Praxis. Symposiumsvorträge, Heidelberg 1996*. Tübingen: Niemeyer, 61–68.
Herberg, Dieter (1998): Auf dem Weg zum deutschen Neologismenwörterbuch. In: Zettersten, Arne/ Pedersen, Viggo Hjørnager/Mogensen, Jens Erik (eds.): *Symposium on Lexicography VIII. Proceedings of the Eighth International Symposium on Lexicography May 2–4,1996, at the University of Copenhagen*. Tübingen: Niemeyer, 187–192.
Kinne, Michael (1989): Endlich: Ein deutsches Neologismenwörterbuch. In: *Der Sprachdienst* 4, 115–116.
Klosa, Annette/Lüngen, Harald (2018): New German Words. Detection and Description. In: Čibej, Jaka/Gorjanc, Vojko/Kosem, Iztok/Krek, Simon (eds.): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana: Ljubljana University Press, Faculty of Arts, 559–569.
Lemnitzer, Lothar (2010): Neologismenlexikographie und das Internet. In: *Lexicographica* 26, 65–78.
Möhrs, Christine (2020): Grübelst du noch oder weißt du es schon? – Glossare erklären Corona-Schlüsselbegriffe. In: *Sprachreport* 3, 30–37.
Weinert, Martin (2021): *Krisensprache – Sprachkrise – Krisenkommunikation. Sprache in Zeiten der COVID-19-Pandemie*. Baden-Baden: Tectum-Verlag.
Wengeler, Martin/Roth, Kersten Sven (eds.) (2020): *Corona. Essayistische Notizen zum Diskurs*. Themenheft aptum. Zeitschrift für Sprachkritik und Sprachkultur 16, 02/03.
Wiegand, Herbert Ernst (1990), Neologismenwörterbücher. In: Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (eds.): *Wörterbücher – Dictionaries – Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Berlin/New York: de Gruyter, 2185–2187.

Kilim Nam, Jinsan An, Hae-Yun Jung

The emergence and spread of Korean COVID-19 neologisms in news articles and user comments and their lexicographic description

1 Introduction

COVID-19-related new words have been coined extensively since December 2019, reflecting linguistic response to a social reality and indicating the dynamics of new words to cope with unprecedented pandemic situations. From an internal linguistic perspective, COVID-19 neologisms are first of all a set of new words which are focused on a specific period and a specific topic and show particular tendencies in terms of grammar and semantics. Not only do they hold great interest in word-formation research, but they also shed light on the relationship between language and discourse communities as they reveal the impact of the pandemic and its perceptions by the public.

There has been much discussion on vocabulary reflecting social and cultural phenomena in relation to lexicographic research. A few decades ago, Williams (1976) published a dictionary of cultural keywords, leading to the development of the 'Keywords Project',[1] which studies diachronic changes of meaning and synchronic meanings of major words. As for Wierzbicka (1999), vocabulary is key to understanding history, culture and society and keywords are evidence that lexicology and lexicography play a central role in interpreting discourse communities.[2]

Positing that the COVID-19 neologisms as a class of vocabulary constitute the keywords of the COVID-19 era, this study aims to analyze the occurrence and usage

1 Details of the project can be found on https://keywords.pitt.edu/ (last access: 10 June 2022).

2 On the other hand, there are also examples of studies on the relationship between certain lexical classes and discourse communities using the quantitative analysis of corpus linguistics, such as Stubbs (1996, 2001) who discussed how collocations and the semantic preferences of certain words could evidence cultural and ideological characteristics, and Scott (2010) who discussed the methodology as well as the importance of keyword extraction based on statistical significance.

Kilim Nam, Kyungpook National University, College of Humanities, Room no.404, 80 Daehak-ro, Daegu, Republic of Korea, e-mail: nki@knu.ac.kr
Jinsan An, Kyungpook National University, Humanities Korea Hall, Room no. 315, 80 Daehak-ro, Daegu, Republic of Korea, e-mail: siveking@naver.com
Hae-Yun Jung, Kyungpook National University, Humanities Korea Hall, Room no. 315, 80 Daehak-ro, Daegu, Republic of Korea, e-mail: haeyun.jung.22@gmail.com

https://doi.org/10.1515/9783110798081-004

of such keywords and provide suggestions for their lexicographic representation, from the perspective of corpus linguistics. While research on the sociolinguistics of neologisms or the correlation between culture and neology has focused hitherto on particular types of neologisms and sociocultural phenomena, the use patterns of neologisms depending on genres and registers has not been fully discussed. However, the proliferation of Web languages and the dynamics of language resulting from mass communication call for the study of neologisms in relation to genres and registers. Thus, this study investigates the neologisms extracted not only from online news articles but also from the comments accompanying such articles. Comments written by non-experts are indeed equally crucial to examine as the articles written by experts since their respective values for analysis differ in terms of production and use of neologisms.

The present paper examines the usage of 341 COVID-19 neologisms which appeared in South Korea over a span of eighteen months (from December 2019 to May 2021) and were extracted from a corpus composed of COVID-19-related news articles and comments, the *COVID-19 Corpus*, in order to address the following research questions: 1) How do the 341 COVID-19 neologisms extracted rank in news articles and comments respectively?, 2) What usage trends do neologisms designating the disease and other high-frequency neologisms show in news articles and comments respectively?, 3) What characteristic differences do comments as a non-expert and subjective language resource and news articles as an expert and objective language resource show and what value may each genre add to the lexicographic description of neologisms?

The following section introduces the composition of the *COVID-19 Corpus* and research methodology. Section 3 provides a quantitative analysis of the COVID-19 neologisms and Section 4 is a case study of the usage trends of the high-frequency neologism *K-pangyek*[3] 'K-quarantine'. Finally, Section 5 discusses a number of issues regarding the lexicographic description of such neologisms as *K-pangyek*.

2 Object of study and methodology

2.1 Extracting COVID-19 neologisms and building the *COVID-19 Corpus*

This paper targets 341 neologisms related to COVID-19, which were coined between December 2019 and May 2021. This list combines the lists of COVID-19 neologisms presented in Lee/Kang/Nam (2020) and Nam et al. (2021), which were expanded with

3 This paper follows the Yale romanization system in transcribing Korean words.

neologisms extracted manually.[4] Lee/Kang/Nam (2020) and Nam et al. (2021) slightly differ in the composition of the corpus used for extraction, the time span for manual extraction, and the identifying criteria.

Table 1: Comparison of the extraction of COVID-19 neologisms.

	Lee/Kang/Nam (2020)	Nam et al. (2021)
methodology	extraction using a domain-general and domain-specific newspaper corpus	extraction using a domain-general newspaper corpus
identifying criteria	– occurrence timescale: December 2019 – August 2020 – occurrence frequency: 1 or more	– occurrence timescale: July 2019 – June 2020 – occurrence frequency: 3 or more
neologisms extracted	302	229

As seen in Table 1, Lee/Kang/Nam (2020) made use of domain-specific newspapers containing the character strings *kholona* 'corona'[5] and *khopitu* 'COVID' in addition to domain-general newspapers and extracted as many neologisms as possible by lowering the minimum occurrence frequency to 1. Nam et al. (2021), on the other hand, collected 405 new words for the year 2020 using only a domain-general newspaper corpus, from which 229 new words directly or indirectly related to COVID-19 were selected and classified as COVID-19 new words. Both studies extracted neologisms from news articles only and similarly excluded controversial slang and discriminatory expressions because the scope of the research was set as an extension of a state-led neologism research project in both cases.

The present study combines the two lists from the above studies, to which were added expressions with pejorative connotations, such as *cwungkwuk pyeylyem* 'Chinese pneumonia', *ccangkkay pyeylyem* 'Chinky pneumonia', *ccangkkay kholona* 'Chinky corona', as well as phrases that appeared after August 2020, including *wulye pyeni* 'fear

4 The extraction of neologisms is carried out in four stages. First, a corpus is built by crawling news articles on the Naver platform (https://www.naver.com/) within a specific time span and a specific scope of media. A list of neologism candidates is compiled by extracting nouns from the corpus based on probability and by extracting inflected and uninflected words via morphological analysis. The list is then reviewed to determine whether the word forms extracted are already listed as headwords in the main online dictionary *Urimalsaem*. The date of first occurrence is checked in Naver News (https://news.naver.com/) and the word forms that meet the identifying criteria are compiled as neologisms. In addition, manual extraction is carried out in parallel for a more comprehensive extraction.

5 Korean uses the abbreviated form *corona* for *coronavirus* to designate the virus and the disease, and often to create COVID-19 related neologisms.

transition’, *pwusuthe syos* ‘booster shot’, *popoksopi* ‘revenge spending’, *payksin cengchi* ‘vaccine politics’, and *pangyek cengchi* ‘quarantine politics’.

Table 2 summarizes the composition of the *COVID-19 Corpus* used to analyze the usage of COVID-19 neologisms.

Table 2: Composition of the *COVID-19 Corpus*.

	Article Corpus	**Comment Corpus**
genre	COVID-19 related news (news articles)	COVID-19 related news (comments on news articles)
time scale	January 1st, 2020 – March 31st, 2021 (15 months in total)	
target	Articles and comments of news pages wherein the word *kholona* (‘corona’ for COVID-19) appears at least 3 times[6]	
number of items	456,680 articles	13,470,427 comments
number of *ecel* (Korean word units)	122,052,029	159,506,286

The *COVID-19 Corpus* is divided into two sub-corpora, one composed of news articles (‘Article corpus’) and the other consisting of the readers’ comments on each article (‘Comment corpus’). The latter is much larger with regard to both the number of items and the amount of text than the former. The articles selected to build the corpus include news articles such as reports on the development of the pandemic (1.a) and articles on the impact of the pandemic on individual lives as well as culture and society (1.b, 1.c).

(1) a. [Breaking news] 465 domestic contacts with confirmed cases, 70 tests performed, 12 confirmed cases of new coronavirus
b. [Headline] Film industry in crisis together with the new coronavirus . . . The release of *The Princess* delayed.
c. Let’s go to the sensible ‘stay at home’ online playground!

While news articles constitute a genre written by experts and characterized by objectivity, readers’ comments form a genre produced by non-expert and characterized by subjectivity, thus containing personal opinions, feelings, and any additional information readers want to share, rather than objective facts. Examples (2) and (3) show

6 Once the corpus was built, some of the COVID-19 neologisms were pre-processed by unifying spellings, removing unnecessary spaces and new lines, and removing special symbols (e.g., *k-quarantine*, *k quarantine*, *K-quarantine*, *K-quarantine*).

typical examples from articles and comments that are concerned with the 'emergency disaster relief fund' (2) and 'corona-blues' (3).

(2) a. The government has announced a policy of 'emergency disaster relief fund' for the bottom 70% of taxpayers.
b. Think about applying the emergency disaster relief fund where it's really needed, instead of giving it out to the whole country.

(3) a. The term 'corona-blues' was even coined in reference to the 'depressed' meaning of 'blue'.
b. Hopefully it will be less tomorrow? It's really corona-blues. . . I'm getting so depressed :(

As can be seen in the above Examples, (2.a) and (3.a), which have been extracted from the 'Article corpus', aim to convey objective information such as reporting on a governmental policy or defining a neologism, whereas the comments (2.b and 3.b) convey the readers' personal opinions and attitudes on the matter.

Up until now, the extraction and description of Korean neologisms have mostly focused on news articles. This study, however, aims to examine comments made by the public too and to compare them to the articles. This would allow us to see how the public's response to the pandemic is expressed in language. Moreover, dictionary descriptions of the COVID-19 neologisms need to represent people's creative language use patterns beyond objective facts such as reports or promotion of national policies.

2.2 Methodology

In order to analyze the usage patterns of our 341 COVID-19 neologisms, we have built both the 'Article corpus' and 'Comment corpus' as monthly corpora, classified the neologisms into a semantic category system, and carried out the usage trend, collocation and n-gram analyses.

The first step is to look into the quantitative characteristics of the 341 neologisms by calculating the overall use frequency of each COVID-19 neologism for each month so as to determine the usage trend of each neologism since its first occurrence. As will be discussed in section 3, it appears that neologisms that have long gone out of use in news articles still appear frequently in the comments, which leads us to question whether it is possible to perform comprehensive frequency analysis for neologisms. The neologisms are classified into twelve semantic categories, including economy, society, education, health, and religion. The analysis of the main semantic domains sheds light on the semantic characteristics of COVID-19 neologisms, which turn out to be different from the characteristics of previous years' neologisms.

The focus is then brought on 35 high-frequency neologisms occurring in both articles and comments, for which quantitative and qualitative analyses are carried out, and the case of the neologism *K-pangyek* 'K-quarantine' in particular is examined to explore the applicability of the results in dictionary descriptions. The case study of *K-quarantine* entails a three-dimensional analysis of the characteristics of the COVID-19 neologisms rather than a simple statistical analysis of the collocates for a particular expression. In other words, we analyze the characteristics of the COVID-19 neologisms by analyzing n-grams and secondary collocates co-occurring with the node-collocate pairs in addition to the primary collocate analysis.

Primary collocates and secondary collocates form a network-based method for collocation analysis as proposed by Brezina et al. (2015), which consists of expanding gradually the existing node A → collocate B relationship into a network of node A → collocate B → collocate C of collocate B and so on.[7] This method allows a better and richer understanding of a text's context.

(4) a. He also expressed his aspiration to build a **new South Korea** by continuing the **success** of **K-quarantine**, which has become a **global standard**.
b. The great **K-quarantine**... A **disgusting** regime that **takes credit for** its **success** by using its people as **guinea pigs**.

K-quarantine and *success* frequently co-occur in both articles (4.a) and comments (4.b). However, it is not possible to pinpoint the usage patterns of *K-pangyek* 'K-quarantine' in the media by solely examining the primary collocate *success*. A closer look on secondary collocates shows that the pair *K-quarantine-success* often co-occur with words such as *global, standard, plan* in articles, while it tends to co-occur with expressions such as *disgusting, take credit for* (all collocates are highlighted in bold in the examples) in comments. Furthermore, by setting the part of speech of secondary collocate to adjectives, the extraction of evaluation and sentiment expressions for a given COVID-19 neologism becomes all the more effective. Accordingly, the extraction criteria for primary and secondary collocates are shown in Table 3.

In addition, we performed an n-gram analysis, i.e. an analysis strings of high-frequency morphemes from 5-grams to 10-grams, to examine the typical context of the neologism and understand its usage context beyond the morpheme and word unit. The n-gram analysis was mainly based on strings with verbs and adjectives as heads.

7 Brezina et al. (2015) also introduce Graphcoll, a visualization tool for collocation analysis. A representative study using Graphcoll is Baker (2016) which explains the network and graph theories making the basis of Graphcoll and discusses the necessity of secondary collocate analysis by comparing the results of the primary collocate analysis for the word *troops* using AntConc and WordSmith with the results yielded by the secondary collocate analysis using Graphcoll. See https://www.futurelearn.com/info/courses/corpus-linguistics/0/steps/104876 (last access: 10 June 2022).

Table 3: Extraction criteria for primary and secondary collocates.

	Co-occurrent target	Co-occurrence window
Primary collocates	common noun (NNG), proper noun (NNP), verb (VV), adjective (VA), adverb (MAG)	±2
Secondary collocates	adjective (VA)	sentence or whole comment

3 Characteristics of high-frequency COVID-19 neologisms and semantic categories

Although this section examines the overall characteristics of the 341 neologisms collected, the focus is on high-frequency neologisms appearing in articles and comments respectively. Table 4 shows the statistical characteristics of the 341 COVID-19 neologisms by semantic categories.[8]

Table 4: Examples of COVID-19 neologisms classified by semantic categories.

Semantic category	Number	Ratio	Examples
Politics and Administration	114	33.4	*kinkupcaynanciwenkum* 'emergency disaster relief fund', *thulaypulpepul* 'travel bubble', *kongcekmasukhu* 'official mask'
Health and Medicine	75	22.0	*pwusuthesyos* 'booster shot', *tolphakamyem* 'breakthrough infection', *keyicintankhithu* 'K-testing kit'
Lifestyle	51	15.5	*kholonaihon* 'corona-divorce', *payksinhyuka* 'vaccine holiday', *cipkhoksitay* 'stuck-at-home era'
Economy	33	9.7	*kholonomisyokhu* 'coro[navirus] [eco]nomy shock', *popoksopi* 'revenge spending', *khyukhonomi* 'Q [uarantine]-[eco]nomy'
Society	23	6.7	*enthaykthusahoy* 'un-[con]tact society', *tekpwuneychayllinci* 'thank you challenge', *caytheylkunmwu* 'working-at-home in hotels'

8 For the discussion on the semantic category system for the COVID-19 neologisms, see Lee/Kang/Nam (2020: 158–161).

Table 4 (continued)

Semantic category	**Number**	**Ratio**	**Examples**
Human beings	13	3.8	*hwakccinja* 'confirmed fatter case', *homomasukwusu* 'homo maskus'
Education	11	3.2	*kholonakheyisyen* 'corona-[va]cation', *enthaykthulening* 'un-[con]tact learning'
Clothing	8	2.3	*epheweye* 'upperwear', *phaynthikhoting* 'panty coding'
Food	6	1.8	*talkonakhephi* 'dalgona coffee', *tolpaptolpap* 'cooking after cooking'
Culture	5	1.5	*enthaykthumwunhwa* 'un-[con]tact culture', *laynsenumakhoy* 'LAN line concert'
Housing	1	0.3	*pitaymyencwukemwunhwa* 'non-face-to-face housing culture'
Religion	1	0.3	*ciphapyeypay* 'group Mass'
Total	341	100	

This semantic category system for neologisms was established in 2015 by the Korean Neologisms Investigation Project, which had been conducted every year from 1994 to 2019 under government supervision (Nam 2019). The classification of neologisms into twelve semantic categories, including society, education, economy, and politics, shows which domains are the most conducive to the creation of neologisms each year and hence each period.

As shown in Table 4, the predominant domains for COVID-19 neologisms are Politics and Administration (33.4%) and Health and Medicine (22%), the two combined accounting for more than half of the neologisms, followed by Life (15.5%), Economy (9.7%), and Society (6.7).[9] In contrast, the predominant semantic categories of 2015, for example, were Society and Economy. Back then, common neologisms related to major economic and social issues, while COVID-19 neologisms are primarily concerned with various welfare policies and health matters. In other words, neologisms have been abundantly created in the fields of politics and medicine as a result of the pandemic crisis.[10]

9 The trends for these semantic categories have been similarly discussed in Lee/Kang/Nam (2020), which targeted 302 COVID-19 neologisms.

10 The proportion of COVID-19 neologisms in the category of Politics and Administration is particularly meaningful considering that this semantic category used to account for only around 5% of the neologisms each year according to the Korean Neologisms Investigation Project.

This is also evident in the high-frequency neologisms listed in Table 5. These neologisms are presented by corpus, ranked from 1 to 35 in order of high frequency, and correspond to 10% of the total 341 COVID-19 neologisms.

All of the top 35 neologisms listed Table 5 are terms either designating the virus, such as *kholona19* 'corona-19', *sincongkholonapailesukamyemcung* 'novel coronavirus disease', and *wuhanpyeylyem* 'Wuhan pneumonia', or referring to welfare measures, such as *sahoycekkelitwuki* 'social distancing', *saynghwalsokkelitwuki* 'ordinary social distancing', *kwukmincaynanciwenkum* 'national disaster relief fund' and *kinkupcaynanciwenkum* 'emergency disaster relief fund'.

In addition, 22 neologisms (highlighted in grey) are common to both news articles and comments. However, depending on whether they are from a news article or a comment, they differ in frequency ranking, usage trends, and the distribution of their collocates (this will be discussed in Section 4 in more detail).

The unmarked neologisms are those which are found in one genre exclusively among its top 35 neologisms, either because they are used only in one genre across the whole sub-corpus or because they made it to the top 35 in one genre only. In the latter case, it is worth examining the reasons why they show such differences in rank and frequency.

For example, *kholonapailesukamyemcung-19* 'coronavirus disease 19' ranks 17th in the Article corpus but only 76th in the Comment corpus; on the other hand, *khopitu-19* 'COVID-19' ranks 23rd in the comments but is down to 58th in articles. These examples show the tendency to use official terms in articles, while shorter forms are preferred in comments. Furthermore, we found a few nicknames to designate the disease in high frequency in the comments (*taykwukholona* 'Daegu corona', *cwungkwukphyeylyem* 'China pneumonia', *cwungkwukkholona* 'China corona', *taykwuphyeylyem* 'Daegu pneumonia', *ccangkkayphyeylyem* 'Chinky pneumonia'), which belong for the most part to hate speech against specific countries or regions (China, Daegu) and even social groups such as Shincheonji members (e.g. *sinchenciphyeylyem* 'Shincheonji pneumonia).[11] Hate and discrimination speech is highly controversial and often banned from newspapers. Although these constitute only a few cases, the use of people's comments rather than the official terms in news articles for language description presents the advantage of showing the raw usage of language. Nonetheless, it also raises the issue of how to assess and deal with users' unrefined language as well as discrimination and hate discourses. If expressions belonging to discrimination or hate speech are to be represented in the dictionary, lexicographers need to assess the necessity of their inclusion in (or exclusion from) dictionaries. But before raising the question of lexicographic inclusion of such terms, lexicographers need to

11 Shincheonji is a religious sect, the members of which are believed to have acted as superspreaders of COVID-19 in February 2020 in Daegu, whereas only a couple of cases had been detected in the entire country at that time.

Table 5: Top 35 COVID-19 neologisms in order of high frequency.

Rank	Articles		Comments	
	Headword	Frequency	Headword	Frequency
1	*kholona19* 'corona 19'	1,810,460	*K-pangyek* 'K-quarantine'	103,695
2	*sincongkholonapailesukamyemcung* 'novel coronavirus disease'	271,123	*wuhanphyeylyem* 'Wuhan pneumonia'	95,117
3	*sahoycekkelitwuki* 'social distancing'	65,650	*kholona19* 'corona 19'	82,704
4	*wuhan phyeylyem* 'Wuhan pneumonia'	22,073	*sahoycekkelitwuki* 'social distancing'	28,593
5	*phosuthukholona* 'post-corona'	18,154	*wuhan kholona* 'Wuhan corona'	16,895
6	*kholonasiday* 'corona era'	17,868	*kholonacengchi* 'corona politics'	6,186
7	*saynghwalchilyoseynthe* 'live-in treatment center'	15,631	*taykwukholona* 'Daegu corona'	5,603
8	*enthaykthu* 'un-[con]tact'	11,484	*cwungkwukphyeylyem* 'China pneumonia'	4,950
9	*leymteysipilu* 'remdesivir'	8,731	*kholonasiday* 'corona era'	4,123
10	*kinkupcaynanciwenkum* 'emergency disaster relief fund'	8,597	*cwungkwuk kholona* 'China corona'	4,016
11	*K-pangyek* 'K-quarantine'	7,920	*theksukhu* 'chin [ma]sk'	3,507
12	*ciphapkumcimyenglyeng* 'gathering restriction'	4,734	*taykwuphyeylyem* 'Daegu pneumonia'	2,665
13	*khohothukyengli* 'cohort isolation'	4,720	*kongcekmasukhu* 'official mask'	2,287
14	*caynankiponsotuk* 'disaster basic income'	4,485	*leymteysipilu* 'remdesivir'	2,275
15	*kholonapullwu* 'corona blues'	4,426	*onlainkayhak* 'online school start'	2,222
16	*saynghwalsokkelitwuki* 'ordinary social distancing'	4,287	*tonghakkaymi* 'patriotic Korean investors'	1,887
17	*kholonapailesukamyemcung-19* 'coronavirus disease 19'	3,310	*khohothukyengli* 'cohort isolation'	1,706
18	*cencachwulipmyengpu* 'electronic entry log'	2,832	*ciphapkumcimyenglyeng* 'gathering restriction'	1,604

19	*tekpwuneychayllinci* ‘thank you challenge’	2,536	*kwukmincaynanciwenkum* ‘national disaster relief fund’	1,563
20	*wuhan kholona* ‘Wuhan corona’	2,376	*kinkupcaynanciwenkum* ‘emergency disaster relief fund’	1,472
21	*kholonapaynteymik* ‘corona pandemic’	2,103	*ccangkkay phyeylyem* ‘Chinky pneumonia’)	1,224
22	*onlainkayhak* ‘online school start’	1,823	*caynankiponsotuk* ‘disaster basic income’	1,191
23	*kinkupkoyongancengciwenkum* ‘emergency employment security relief fund’	1,799	*khopitu19* ‘COVID-19’	1,097
24	*onthaykthu* ‘on[line] [con]tact’	1,655	*saynghwalsokkelitwuki* ‘ordinary social distancing’	995
25	*kongcekmasukhu* ‘official mask’	1,643	*phinseyspangyek* ‘tweezer quarantine’	924
26	*tonghakkaymi* ‘patriotic Korean investors’	1,635	*enthaykthu* ‘un-[con]tact’	873
27	*wekhingsulwu* ‘walking-through’	1,317	*kholonapullwu* ‘corona blues’	853
28	*kwukmincaynanciwenkum* ‘national disaster relief fund’	1,121	*sincongkholonapailesukamyemcung* ‘novel coronavirus disease’	817
29	*pimalchatan* ‘cough shield’	1,072	*kkamkkamikamyem* ‘mysterious infection’	786
30	*kikansanepanjengkikum* ‘primary industry stability fund’	1,064	*phosuthukholona* ‘post-corona’	719
31	*witukholona* ‘with corona’	1,041	*kholonacencayng* ‘corona war’	702
32	*teyksathason* ‘dexamethasone’	975	*khosukhu* ‘nose [ma]sk’	676
33	*saynghwalbangyekswuchik* ‘daily life quarantine guidelines’	902	*saynghwalchilyoseynthe* ‘live-in treatment center’	651
34	*tulaipusulwusenpyelcinlyoso* ‘drive-through screening station’	887	*kholonapaynteymik* ‘corona pandemic’	582
35	*cateyktayphi* ‘home-as-shelter’	876	*kholonathukswu* ‘corona special’	485

define the boundaries between discrimination and hate expressions from a legal perspective and from a linguistic perspective.

Some technical terms that are highly frequent in the Article corpus, such as *onthaykthu* 'on[line] [con]tact' or welfare policy names (*kinkupkoyongancengciwenkum* 'emergency employment security fund', *kikansanepanjengkikum* ' . . . industry stability fund'), are ranked quite low in the Comment corpus (80th, 156th and 151st respectively), while blend words, such as *theksukhu* 'chin [ma]sk' or *khosukhu* 'nose [ma] sk', show the opposite tendency, ranking 11th and 32nd respectively in the comments but only 65th and 138th respectively in the articles. Another case in point is *kholonacengchi* 'corona politics', which is highly frequent in comments (6th) compared to articles (119th). *Kholonacengchi* 'corona politics' is a sarcastic term to designate the use of COVID-19 for political purposes. These examples show that in comments, not only are creative neologisms more commonly used but they also tend to convey language users' opinions and assessments of the COVID-19 situation.

It is worth noting that the same neologism may also exhibit different characteristics in terms of usage depending on whether it is used in news articles or in the comments. For instance, in the articles, *theksukhu* 'chin [ma]sk' and *khosukhu* 'nose [ma]sk' are used when defining the neologism and in reference to related policies and measure promotion (5), whereas in the comments, the terms are used to express one's attitudes, opinions and/or feelings regarding *theksukhu*- and *khosukhu*-related phenomena (6).

(5) a. 'Chin-mask' and 'nose-mask' still out there, further action needed beyond regulations.
b. The government has warned against the so-called 'nose-mask', whereby people wear masks with their noses exposed.
c. Mask or no mask, Seoul cracks down on 'chin-mask' and 'nose-mask'.

(6) a. I want to punch everyone who wears chin-mask and nose-mask.
b. People wearing chin-mask, nose-mask or no mask, behave yourself.
c. I really hate seeing people wearing chin-mask or nose-mask these days.

These examples show that the usage characteristics of neologisms are obtained all the more clearly with concordance and collocate analysis. As the case of *theksukhu* and *khosukhu* can be extrapolated to many other neologisms, it becomes apparent that language resources such as comments may provide useful data for lexicography research as they show aspects of usage in real-life contexts. This is all the more crucial in the case of neologisms because their dynamicity requires to bring the focus on actual usage.

As discussed above, when handling language data such as people's comments in addition to news articles, it is necessary to adopt a comprehensive approach to the extraction and description of COVID-19 neologisms by considering changes in

the list of headword candidates, different frequency rankings and different usage characteristics according to varying frequencies across genres. In line with such considerations, the following section provides a detailed, qualitative study of *K-pangyek* ‘K-quarantine’.

4 Case study: *K-pangyek* ‘K-quarantine’

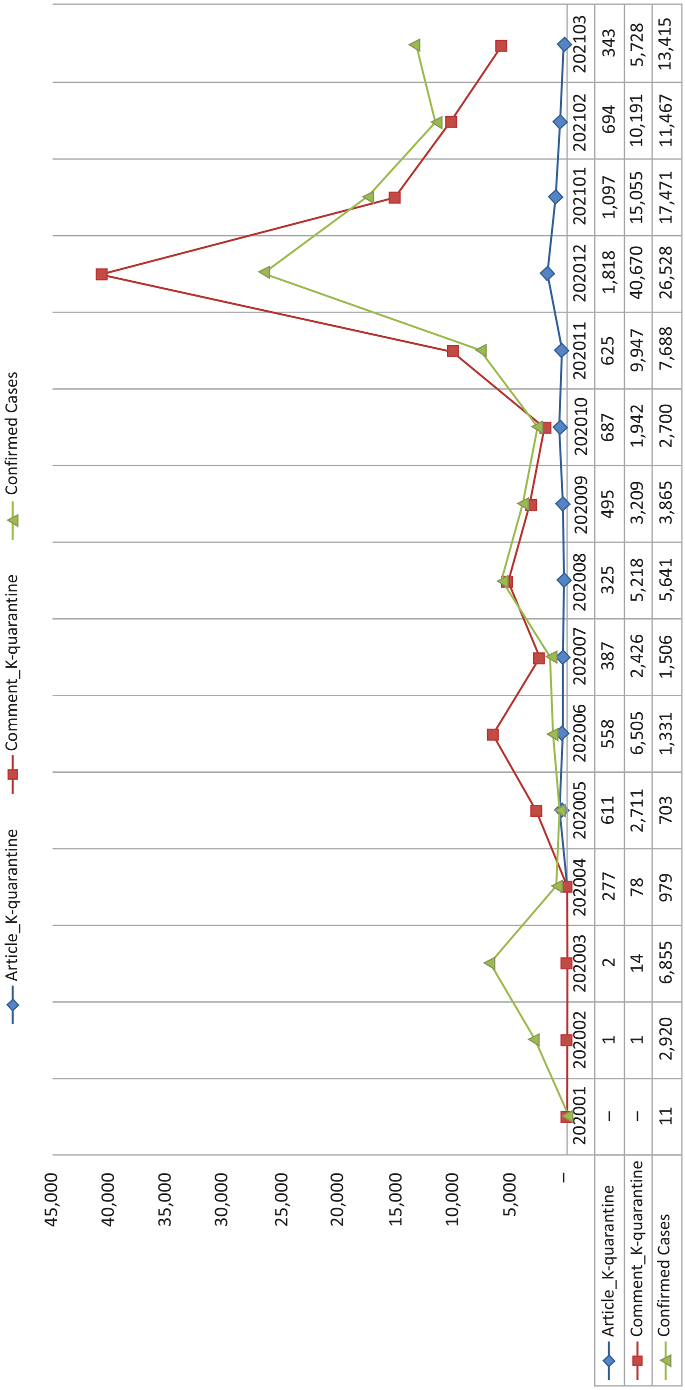
4.1 Primary collocates of *K-pangyek*

K-pangyek ‘K-Quarantine’ ranks 11th in high frequency for the Article corpus and is the most frequently used neologism in the Comment corpus. The term first appeared in February 2020 and was coined in relation to ways of dealing with the COVID-19 outbreak. The letter ‘K’ refers to Korea (i.e., South Korea’s quarantine system). In the early stages of the pandemic, while most countries struggled with quarantine, South Korea established an effective quarantine system using a ‘drive-through/walk-through screening stations’ and ‘live-in treatment centers’, for which it was considered a role model for the international community. However, the evaluation of the K-quarantine may vary greatly depending on the timing of evaluation, but above all, on the medium where it is evaluated, sometimes resulting in contradictory assessments. Indeed, factors such as the social class of a given newspaper’s readership and the speaker’s political orientation may exert great influence.

From a quantitative perspective, *K-pangyek* occurred 7,920 times in the Article corpus and 103,659 times in the Comment corpus from January 2020 to March 2021. Monthly use trends are shown in Figure 1.

The frequency trends of *K-pangyek* show a very strong correlation with the trends in the number of confirmed cases of COVID-19. Summer 2020 began with the second wave of the pandemic in South Korea, the number of infected persons increasing throughout the summer and even doubling from May to June 2020. In parallel, the neologism hit peaks in June and August 2020. The highest peak (for both the neologism and the confirmed cases) was in December 2020, which coincides with the third wave. These trends manifest the failure of the K-quarantine model, fueling fierce criticisms in comments. The third wave, in particular, sparked off strong criticisms on the delay in supplies and inoculations of the COVID-19 vaccine compared to other countries.

To understand the differences in the overall usage of *K-pangyek*, we extracted the 30 most frequent primary collocates occurring within an L2R2 window from both article and comment examples, targeting only content words. Thus, these collocates include common nouns (NNG), proper nouns (NNP), verbs (VV), adjectives (VA), and adverbs (MAG), which are presented in Table 6 in both absolute frequencies and normalized frequencies (1 per 100,000 *ecel*).



	202001	202002	202003	202004	202005	202006	202007	202008	202009	202010	202011	202012	202101	202102	202103
Article_K-quarantine	–	1	2	277	611	558	387	325	495	687	625	1,818	1,097	694	343
Comment_K-quarantine	–	1	14	78	2,711	6,505	2,426	5,218	3,209	1,942	9,947	40,670	15,055	10,191	5,728
Confirmed Cases	11	2,920	6,855	979	703	1,331	1,506	5,641	3,865	2,700	7,688	26,528	17,471	11,467	13,415

Figure 1: Frequency of *K-pangyek* ‘K-quarantine’ per month.

Table 6: Top 30 primary collocates of *K-pangyek*.

Rank	Articles				Comments			
	Morpheme	Tag	Abs. fr.	Norm. fr.	Morpheme	Tag	Abs. fr.	Norm. fr.
1	*sengkwa* 'achievement'	NNG	473	334	*ha* 'do'	VV	10768	475
2	*seykyey* 'world'	NNG	454	321	ㅋ 'haha'	NNG	9596	424
3	*hongpo* 'promotion'	NNG	447	316	*calang* 'boast'	NNG	3499	154
4	*sengkong* 'success'	NNG	359	254	*seykyey* 'world'	NNG	3338	147
5	*cengpwu* 'government'	NNG	357	252	*hongpo* 'promotion'	NNG	3293	145
6	*kwukmin* 'citizen'	NNG	313	221	*calangha* 'boast'	VV	3218	142
7	*cheykhu* 'check'	NNG	290	205	*kwukmin* 'citizen'	NNG	3041	134
8	*phaykthu* 'fact'	NNG	289	204	*cengpwu* 'government'	NNG	2760	122
9	*moteyl* 'model'	NNG	285	201	*mwunjayin* 'Moon Jae-in'	NNP	2708	120
10	*kholona-19* 'COVID-19'	NNG	265	187	*cahwacachanha* 'self-praise'	VV	2568	113
11	*hankwuk* 'South Korea'	NNP	228	161	*pangyek* 'prevention'	NNG	2516	111
12	*toy* 'become'	VV	223	158	*eps* 'not be'	VA	2457	108
13	*is* 'link'	VV	205	145	*payksin* 'vaccine'	NNG	2424	107
14	*kwukcey* 'international'	NNG	195	138	*calha* 'do well'	VV	2346	104
15	*tayha* 'about'	VV	184	130	*iss* 'be'	VA	2282	101
16	*mopem* 'role model'	NNG	177	125	*cahwacachan* 'self-praise'	NNG	2143	95
17	*wuswuseng* 'excellence'	NNG	176	124	*silphay* 'failure'	NNG	2090	92
18	*pangyek* 'prevention'	NNG	174	123	*toy* 'become'	VV	2054	91
19	*ha* 'do'	VV	172	122	*ka* 'go'	VV	2052	91
20	*cen* 'whole'	NNG	163	115	*kath* 'similar'	VA	1971	87
21	*pyocwun* 'standard'	NNG	149	105	*silchey* 'true colour'	NNG	1953	86

Table 6 (continued)

Rank	Articles				Comments			
	Morpheme	**Tag**	**Abs. fr.**	**Norm. fr.**	**Morpheme**	**Tag**	**Abs. fr.**	**Norm. fr.**
22	*kyenghyem* 'experience'	NNG	148	105	*kholona* 'corona'	NNG	1783	79
23	*pat* 'receive'	VV	147	104	*ttay* 'when'	NNG	1615	71
24	*ceyphwum* 'product'	NNG	144	102	*ecce* 'how'	VV	1610	71
25	*wiki* 'crisis'	NNG	143	101	*ttetul* 'make noise'	VV	1594	70
26	*taythonglyeng* 'president'	NNG	139	98	*an* 'not'	MAG	1593	70
27	*calangha* 'boast'	VV	137	97	*choyko* 'best'	NNG	1483	65
28	*kholona* 'corona'	NNG	136	96	*honpoha* 'promote'	VV	1465	65
29	*tto* 'too'	MAG	124	88	*sengkong* 'success'	NNG	1462	65
30	*eps* 'not be'	VA	119	84	*soli* 'noise'	NNG	1457	64

Exclusive collocates that appear as primary collocates only in one genre have been highlighted in grey. These are particularly useful in showing the various evaluation tendencies in articles and comments. In the case of articles, exclusive keywords, such as *sengkwa* 'achievement', *moteyl* 'model', *wuswuseng* 'excellence', or *pyocwun* 'standard', show a positive evaluation of K-quarantine, while in the comments, evaluations appear rather negative with frequent collocates such as *cahwacachanha* 'self-praise', *silphay* 'failure', *silchey* 'true colour', and *ttetul* 'make noise'. As can be seen in the examples below, exclusive collocates are representative of the evaluation of the *K-quarantine* made either in articles or in comments. Example (7) shows the typical case of the collocate *moteyl* 'model' in articles, and (8) the typical case of the collocate 'Moon Jae-in'[12] in comments.

(7) a. The K-quarantine model as a reliable shield against COVID-19, has been shared with the whole world.
b. After achieving global recognition with its drive-through screening stations and testing kits, the K-quarantine model has also yielded impressive results.

12 Examples with the collocate 'Moon Jae-in' are generally 'negative evaluation'; however, there are a few cases of positive evaluation, such as: Thumbs for President Moon Jae-in's 'K-quarantine' that is recognized around the world! All Koreans respect it and support it.

c. Singapore introduces the ‘drive-through screening station’ of the K-quarantine model.

(8) a. K-quarantine has failed because of Moon Jae-in.
b. Moon Jae-in called every country to brag about sharing the secret of K-quarantine but look what it’s become now.
c. Moon Jae-in has spent 120 billion won to promote a K-quarantine that does not even work. And then he had no more money to buy vaccines.

In addition to the exclusive collocates, the top 30 most frequent collocates also include common collocates, such as ‘*cengpwu* ‘government’, *sengkong* ‘success’, or *kholona* ‘corona’, and most likely, other exclusive collocates that did not rank in the top 30. Therefore, a clear difference between articles and comments cannot be asserted categorically with the sole primary collocates presented above. This requires instead additional quantitative analysis, including in particular a primary collocate analysis with a wider window. Moreover, the narrow L2R2 window suffers from the shortcoming that adjectives (VA), which directly express evaluations and emotions, are left out of the primary collocation lists. Therefore, it is necessary to carry out a secondary collocate analysis as well as an n-gram analysis to cope with such shortcomings. The next section first explores the co-occurrences of two common primary collocates of *K-pangyek*, namely *cengpwu* ‘government’ and *payksin* ‘vaccine’ (which has not made it to the top 30 in the case of articles), by analyzing entire sentences and/or comments, and additionally performs a 5-gram analysis. Figure 2 presents the analytic process for co-occurrences of *K-pangyek*, focusing on the two cases of *cengpwu* and *payksin*.

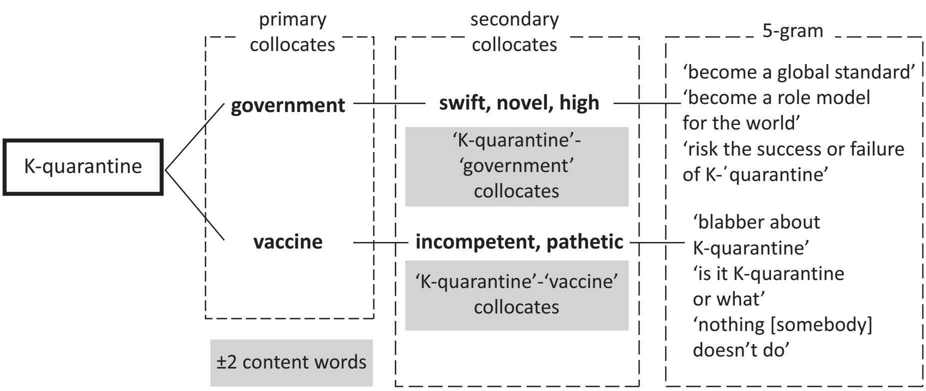


Figure 2: Analytic process for the co-occurrences of *K-pangyek*.

4.2 Secondary collocates and n-grams for *K-quarantine*

4.2.1 *Cengpwu* 'government'

First, let's examine the secondary collocates for the *K-pangyek-cengpwu* pair. These collocates target all the adjectives (VA) in sentences where *K-pangyek* and *cengpwu* are co-occurring, the extraction scope being extended to the whole sentence or comment. Table 7 presents the top 20 secondary collocates in order of high frequency, showing both absolute frequencies and normalized frequencies (1 per 10,000 *ecel*).

Table 7: Top 20 secondary collocates/*cengpwu* 'government' (NNG).

Rank	Articles				Comments			
	Morpheme	Tag	Abs. fr.	Norm. fr.	Morpheme	Tag	Abs. fr.	Norm. fr.
1	*iss* 'be'	VA	192	75	*eps* 'not be'	VA	4714	89
2	*eps* 'not be'	VA	94	37	*iss* 'be'	VA	3416	64
3	*manh* 'many'	VA	44	17	*kath* 'similar'	VA	1727	33
4	*ppalu* 'swift'	VA	33	13	*mwunungha* 'incompetent'	VA	1268	24
5	*kath* 'similar'	VA	29	11	*kuleh* 'like that'	VA	992	19
6	*khu* 'big'	VA	26	10	*ileh* 'like this'	VA	852	16
7	*saylop* 'novel'	VA	20	8	*coh* 'good'	VA	663	13
8	*elyep* 'difficult'	VA	20	8	*manh* 'many'	VA	624	12
9	*nop* 'high'	VA	20	8	*hansimha* 'pathetic'	VA	462	9
10	*kanungha* 'possible'	VA	20	8	*ecceh* 'how'	VA	430	8
11	*pwucokha* 'insufficient'	VA	19	7	*himtul* 'strenuous'	VA	422	8
12	*kuleh* 'like that'	VA	19	7	*khu* 'big'	VA	390	7
13	*sinsokha* 'prompt'	VA	18	7	*nuc* 'late'	VA	323	6
14	*coh* 'good'	VA	17	7	*cwungyoha* 'important'	VA	299	6
15	*cwungyoha* 'important'	VA	16	6	*taytanha* 'impressive'	VA	268	5
16	*phylyoha* 'necessary'	VA	15	6	*cek* 'few'	VA	242	5
17	*ecceh* 'how'	VA	14	6	*kanungha* 'possible'	VA	222	4

Table 7 (continued)

Rank	Articles				Comments			
	Morpheme	Tag	Abs. fr.	Norm. fr.	Morpheme	Tag	Abs. fr.	Norm. fr.
18	*ileh* 'like this'	VA	13	5	*calna* 'distinguished'	VA	207	4
19	*wanpyekha* 'perfect'	VA	13	5	*silh* 'dislike'	VA	196	4
20	*chelceha* 'thorough'	VA	12	5	*phylyoha* 'necessary'	VA	193	4

The exclusive secondary collocates (highlighted in grey) presented in Table 7 show a clearer article-comment divide in the evaluation of the K-quarantine. Articles tend to use more positive adjectives, such as *ppalu* 'swift', *saylop* 'novel', *nop* 'high', *wanpyekha* 'perfect' or *chelceha* 'thorough', while comments tend to contain adjectives conveying rather negative evaluation, including *mwunungha* 'incompetent', *hansimha* 'pathetic', *himtul* 'strenuous', *nuc* 'late', and *silh* 'dislike'. *Mwunungha* 'incompetent' in particular, which is the most frequent exclusive secondary collocate in comments, occurs only once as secondary collocate in articles. Moreover, in the case of comments, other adjectives that have a high frequency but are not in the top 20 include *taptapha* 'frustrating' (171 occurrences), *yekkyepha* 'disgusting' (134 occurrences), *anilha* 'complacent' (132 occurrences), and *mengchengha* 'stupid' (132 occurrences), thus showing a strong tendency to more direct and adverse criticisms in comments as compared to articles.

4.2.2 *Payksin* 'vaccine'

As a primary collocate of *K-pangyek*, *payksin* 'vaccine' ranks 13th in the Comment corpus and 34th in the Article corpus. For the secondary collocates for the *K-pangyek-payksin* pair, we again focused on adjectives (VA) and listed the most frequent ones in Table 8. Just as in Table 7, frequencies are presented in descending order, and in both absolute and normalized (1 per 10,000 *ecel*) values.

Similarly to the case of *cengpwu* and rather unsurprisingly, the exclusive secondary collocates are dominated by positive evaluation in articles (e.g. *thanthanha* 'strong', *sinsokha* 'prompt', *kakkap* 'close') and negative evaluation in comments (e.g. *mwunungha* 'incompetent, *nuc* 'late', *hansimha* 'pathetic'). There are also a few positive adjectives in comments, such as *coh* 'good', *taytanha* 'impressive', and *calangsulep* 'proud'; however, they are often used ironically or sarcastically to mock K-quarantine or criticize the government's policies, as shown in the examples below (the collocates are highlighted in bold characters).

Table 8: Top 20 secondary collocates/*payksin* 'vaccine' (NNG).

Rank	Articles				Comments			
	Morpheme	**Tag**	**Abs. fr.**	**Norm. fr.**	**Morpheme**	**Tag**	**Abs. fr.**	**Norm. fr.**
1	*iss* 'be'	VA	213	109	*eps* 'not be'	VA	5385	92
2	*eps* 'not be'	VA	78	40	*iss* 'be'	VA	3749	64
3	*kath* 'similar'	VA	31	16	*kath* 'similar'	VA	1671	28
4	*kanung* 'possible'	VA	31	16	*mwunungha* 'incompetent'	VA	1160	20
5	*cwungyoha* 'important'	VA	29	15	*kuleh* 'like that'	VA	1002	17
6	*thanthanha* 'strong'	VA	29	15	*nuc* 'late'	VA	836	14
7	*sinsokha* 'prompt'	VA	18	9	*coh* 'good'	VA	820	14
8	*khu* 'big'	VA	18	9	*ileh* 'like this'	VA	700	12
9	*kakkap* 'close'	VA	17	9	*manh* 'many'	VA	588	10
10	*phylyoha* 'necessary'	VA	16	8	*ecceh* 'how'	VA	490	8
11	*mwumoha* 'reckless'	VA	16	8	*hansimha* 'pathetic'	VA	448	8
12	*manh* 'many'	VA	15	8	*cwungyoha* 'important'	VA	420	7
13	*ppalu* 'fast'	VA	14	7	*taytanha* 'impressive'	VA	326	6
14	*wuswuha* 'outstanding'	VA	14	7	*kanungha* 'possible'	VA	310	5
15	*kuleh* 'like that'	VA	13	7	*ppalu* 'fast'	VA	304	5
16	*talu* 'different'	VA	13	7	*himtul* 'strenuous'	VA	301	5
17	*thunthunha* 'robust'	VA	12	6	*calna* 'distinguished'	VA	293	5
18	*nop* 'high'	VA	11	6	*khu* 'big'	VA	265	5
19	*saylop* 'novel'	VA	11	6	*calangsulep* 'proud'	VA	257	4
20	*ancenha* 'safe'	VA	11	6	*phylyoha* 'necessary'	VA	214	4

(9) a. They made a big fuss about **K-quarantine** last year, but now ~ **good** job with the **vaccine**.
b. Bragging sounds **good** . . . the **good** stuff is all about **K-quarantine** . . . but what's the point, **vaccine** inoculations haven't started yet

(10) a. hey bragged about **K-quarantine** crowing that they will be first in the world at developing the **vaccine**, but are we even in the early stages of clinical trials? That's **impressive**. Is this media manipulation?
b. What an **impressive K-quarantine**. Spending 120 billion won in promoting that **K-quarantine**. If they had lobbied for the **vaccine** with that money instead, we would be overflowing with vaccines already. Where on earth did that money disappear? Who the hell used it all?

(11) a. **Proud K-quarantine**: 0 Pfizer **vaccines** secured.
b. Even developing countries have bought Pfizer and Moderna **vaccines** . . . I'm so **proud** of you, damn **K-quarantine**.

While *coh* 'good' did count a few positive evaluations, negative contexts, including sarcastic expressions (9.a: 'good job'; 9.b: 'sounds good'), accounted for most of the cases. In the same vein, more than 90% of the 326 examples containing *taytanha* 'impressive' and the 257 examples containing *calangsulep* 'proud' in the comments are sarcasms, as can be seen in (10) and (11). Commenters make use of irony to criticize the government's emphasis on the K-quarantine success at the beginning of the pandemic, express their disbelief in the government, and convey their anger or frustration about the mismanagement of the vaccine supply and inoculations.

4.2.3 N-gram analysis

The usage examples of 5-grams including *K-pangyek* illustrate the patterns used in high frequency in articles and comments in a more direct manner. Table 9 presents the top 15 5-grams in descending order of high frequency in each sub-corpus.[13]

In the case of articles, the n-grams are mostly extracted from public statements on COVID-19 by a small number of experts, such as spokespersons of the government, political parties, and quarantine authorities, and therefore, contain rather positive evaluations on K-quarantine (n-grams 1, 2, 6, 7, 9, 13, and 14 in the 'Articles' column). On the other hand, comments are produced by a wide breadth of users and exhibit the typical patterns produced by commenters speaking about K-quarantine.

13 5-grams from identical comments posted by the same users on multiple newspapers and 5-grams that are actually parts of 6-grams and above were excluded from the list.

Table 9: Top 15 5-grams for *K-pangyek* 'K-quarantine'.

Rank	Articles		Comments	
	n-grams	**freq.**	**n-grams**	**freq.**
1	*seykyey-uy pyocwun-i toy* 'become a global standard'	90	*K-pangyek kath-un soli ha* 'blabber about K-quarantine'	387
2	*seykyey-uy mopem-i toy* 'become a role model for the world'	74	*K-pangyek-ini mwe-ni ha* 'is it K-quarantine or what'	160
3	*K-pangyek-uy sengpay-lul kel* 'risk the success or failure of K-quarantine'	51	*ha-nun kes-i eps* 'there is nothing [somebody] doesn't do'	142
4	*ek wen-ulo tayphok nulli* 'drastically increase by [number] won'	39	*K-pangyek calang-cil-man ha* 'only brag about K-quarantine'	128
5	*payksin coki hwakpo-ey silpayha* 'fail to secure vaccines early'	38	*un ancwung-ey-to eps* 'not care at all'	102
6	*sengkwa-ka kyengcey-lo ieci* 'achievements are linked to the economy'	37	*K-pangyek-un eti-lo ka* 'what is going on with K-quarantine'	95
7	*tayhanminkwuk-uy cabwusim-i toy* 'be the pride of South Korea'	37	*ha-l ttay-pwuthe alapo* 'know from the moment when'	87
8	*kyenghem-ul seykyey-wa kongyuha* 'share the experience with the world'	36	*1-nyen-i ta toy* 'it's been almost a year'	87
9	*wuswuha-m-ul yesilhi poi* 'clearly show excellence'	35	*masukhu cal ssu-ko tani* 'go around with the mask on'	77
10	*ka toy-l swu iss* 'could become'	34	*so ilh-ko oyyangkan kochi* 'lock the stable after the horse has bolted'	73
11	*cwuchey-ka toy-e ilwu* 'achieve as an agent'	33	*an ha-ko mwe ha* 'what is [somebody] doing, instead of'	67
12	*K-pangyek-uy kipan-i toy* 'become the foundation of K-quarantine'	30	*man ha-l cwul al* 'only know how to'	62
13	*wiki swunkan-ey tewuk kangha* 'stronger in times of crisis'	29	*pyengsang hwakpo-to an ha* 'don't even secure sickbeds'	53
14	*uy sengkong-ul tewuk tuntunha* '[verb] more reliably the success of'	28	*hongpo-pi-lo-man thangcinha* 'squander only on promotional expenses'	53
15	*cinanhay K-pangyek cahwacachan-ey tochwiha* 'intoxicated with the self-praise of last year's K-quarantine'	28	*ha-l mal-i-eps* 'be lost for words'	52

Quite significantly, many of these patterns are composed of high frequency verbs such as *ha* 'do' (n-grams 1, 2, 3, 4, 11, 12, 13, and 15 in the 'Comments' column) and constructions with the adjectives *iss* 'be' and *eps* 'not be' (n-grams 3, 5, and 15), which cannot be grasped with the sole collocate analysis. A sample of corresponding *n-gram* concordance lines for articles and comments are shown in examples (12) and (13) respectively.

(12) N-gram concordances in 'Articles'

a. ***K-pangyek**un **seykyeyuy phyocwuni toy**essko seykyeyeyse kacang ppalli kyengceylul hoypokhako isssupnita*. '**K-quarantine** has **become a global standard** and is rebuilding the economy faster than anything else in the world'

b. *mwun taythonglyengto onul(28il) siceng yenseleyse "**K-pangyek**un cen **seykyeyuy mopemi toy**mye, tayhanminkwukuy capwusimi toyessta." lako phyengkahaysssupnita*. 'Today, the 28th, President Moon has stated in a governmental address that "**K-quarantine is a role model for the** whole **world** and makes the pride of South Korea."'

c. *ichelem nollawul cengtolo palcenhan poken uylyo cheykyeywa paio uyyakphwum sayngsan nunglyeki **K-pangyekuy kipani toy**esssupnita*. 'Such remarkably advanced healthcare system and biopharmaceutical production capacities **were the foundation of K-quarantine**.'

d. ***K-pangyek** yeysanul 1co 8chenek **wenulo tayphok nully**esssupnita*. '**K-quarantine** budget has been **drastically increased to** 1.8 trillion **won**.'

(13) N-gram concordances in 'Comments'

a. *taymanun cikum kholona hwanca = 0 iketun. mwusun **K-pangyek kathun soli hako iss**nya?* 'Right now there is 0 COVID case in Taiwan. What **K-quarantine** are you **blabbering about**?'

b. *ikesi palo mwunceyini calanghanun **K-pangyek**ita. mwe hana ceytaylo **hanun kesi eps**ta*. 'This is the **K-quarantine** Moon Jae-in is boasting about. **There is not one thing he is doing** right.'

c. *cengpwuka kwukminuy kosayng**un ancwungeyto eps**nun kes kathta. wulinalanun wancenhi papo hokwuta. cen seykyey kamyemcatuli mollyeokeyssta. ikey **K-pangyek**ilanta*. 'The government **doesn't** seem to **care about** its citizens' suffering. Our country is a complete laughing stock. People are coming from all over the world. This is what they call **K-quarantine**.'

d. *ikesi **K-pangyek**!! omanpangca**hal ttaypwuthe alapwa**ssta* ㅋㅋㅋ 'This is **K-quarantine!! I knew it from the moment** he started to brag about it hahaha'

e. ***pyengsang hwakpoto anha**y~ payksin hwakpoto anhay~ yethay 1nyentongan cincca han key mweeyyo?? cengpwunimtul?? ceyka pokieyn **K-pangyek** man calangkeli han ke malkon amwukesto mos han ke kathuntey*... 'You

> **didn't even secure sickbeds**~ You don't even secure vaccines~ what is it that you have done for the last one year? Government people?? All I can see is you bragging about **K-quarantine** but nothing else . . .'

N-gram analysis shows the contextual meaning that cannot be analyzed in collocation analysis by extracting patterns from entire sentence or comment as a whole, instead of keeping to the limited vocabulary surrounding *K-Quarantine*. It also confirms the usage trend differences between articles and comments.

L2R2 window primary collocates, secondary collocates, and n-gram analyses of *K-pangyek* demonstrated that despite the fact that the neologism was used with high frequency in both news articles and comments, usage patterns were rather different, if not opposite to each other. Articles tended to praise K-quarantine or at least talk about it in neutral terms as they report on the official stance of the government. Comments, on the other hand, left room for personal opinions, criticisms and negative evaluations of the government's actions, often using irony. If comments cannot be said to represent the entirety of language perfectly, they do constitute 'real-life examples' of subjective evaluation and creative language which cannot be shown in article texts. In that sense, it has become more than necessary to take them into consideration in neologism research and dictionary description.

5 From experts' language to people's language: Suggestions for the lexicographic description of COVID-19 neologisms

As for conclusions, this last section examines ways to incorporate the above analyses into the lexicographic description of COVID-19 neologisms. Research on neologisms has mainly focused on the genre of 'news articles' so far. In the same vein, neologism headwords and usage examples provided by dictionaries for such headwords are often newspaper-based. As demonstrated in the previous sections, comments as language resources by non-expert writers provide a new, raw facet of language that reflects to great extent, at least in terms of neologisms, the creativity and diversity of language users.

The utilization of user-generated content such as comments in neologism description entails a change in the scope and the method of the lexicographic description of neologisms. Here we discuss the case of COVID-19 neologisms in terms of both macrostructure and microstructure.

First of all, it is rather evident that the macrostructure changes according to the language resources used to collect headwords. On the one hand, the example of names designating COVID-19 evidences the potential diversity of media and genres

in neology research, but on the other hand, it shows some limitations and raises issues to be addressed. For instance, the official name of the disease according to the South Korean government is *kholonapailesukamyemcung-19* 'coronavirus infection 19'. Figure 3 shows that only three other names are used more frequently in news articles. In comments, however, the official name is rarely used and instead, there are no less than ten other names which are used more frequently as shown in Figure 4. For the most part, appellations found in comments are strongly biased, thereby quasi-inexistent in articles with the exception of *wuhanphyeylyem* 'Wuhan pneumonia'.[14]

Metcalf (2002) and Barnhart (2007) have proposed the diversity of genres as one of the determining criteria for the establishment of neologisms and their inclusion in the dictionary, along with frequency and time span. Thus, the use of comments as language resource for the extraction of neologisms to be included in Korean dictionaries would mean securing the diversity of the genre where neologisms may potentially appear, but also redefining the other determining criteria, including frequency, which have been traditionally used in Korean neology and lexicography research (Nam 2015). Furthermore, as illustrated in (14), commenters' language, and broadly speaking social media language, more often than not contains socially problematic discriminatory and hate expressions, thereby requiring language experts' examination and discussion with regard to the representation of such deviant language. Regardless, the diversity and non-normative deviance of comments offer great prospects for neologism research.

Korean neologism research has tended to focus on frequency when it comes to the establishment of new words, but comments often show a completely different tendency from articles. A case in point is the neologism *wuhanphyeylyem* 'Wuhan pneumonia' (Lee/Kang/Nam 2020: 154). The neologism was used extensively at the beginning of the pandemic but appeared to fall out of usage as early as March 2020. Indeed, as shown in Figure 5, solely based on the frequency distribution of news articles, *wuhanphyeylyem* is more of a dying neologism; however, the frequency trends of *wuhanphyeylyem* in comments tell a different story, where the neologism is still in usage and has the potential to survive for many more years.

Regarding the microstructure of COVID-19 neologism headwords, comments provide examples of language spoken in real life in contrast with the traditional examples taken from articles, which do not necessarily provide typical usages of neologisms. In addition, the inclusion of socio-cultural characteristics, which can be based on comment data, and contextual and pragmatic information on the headword may be crucial depending on the type and purpose of the dictionary. The following two examples present a microstructural model for the two COVID-19 neologisms of *wuhanphyeylyem*

---

14 In the beginning, newspapers also called the disease *wuhanphyeylyem* 'Wuhan pneumonia' in the sense that it first appeared in Wuhan, China, and manifested as a pneumonia.

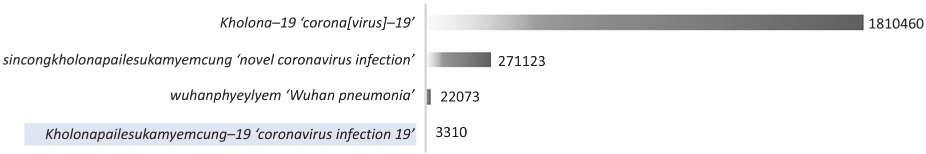


Figure 3: Frequencies of COVID-19 appellations in articles (from January 2020 to March 2021).

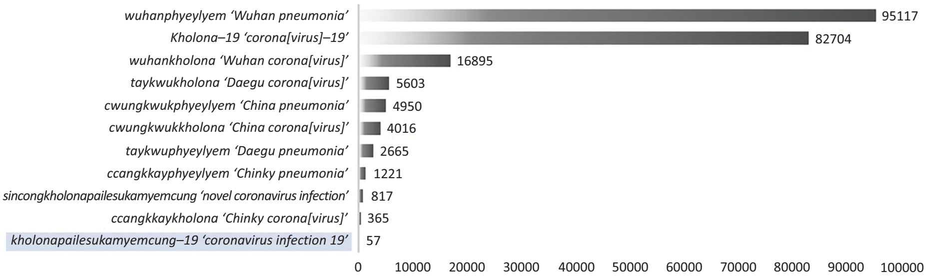


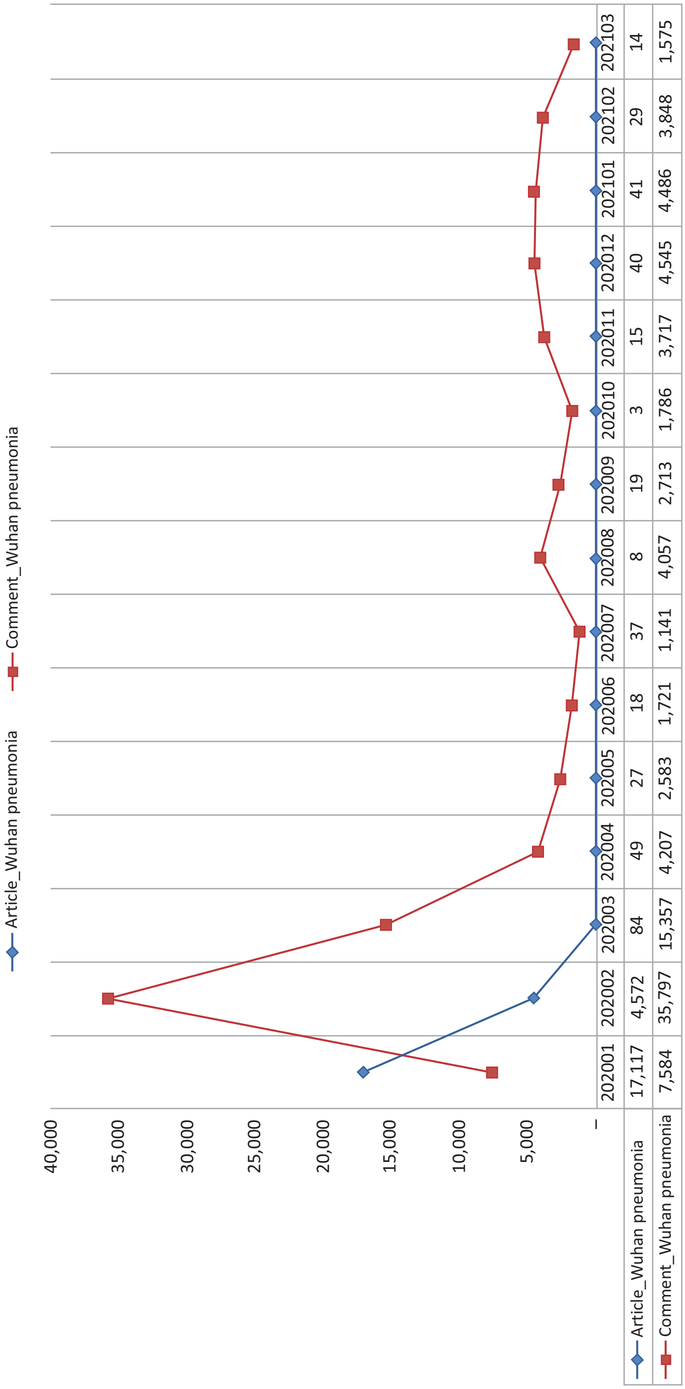
Figure 4: Frequencies of COVID-19 appellations in comments (from January 2020 to March 2021).

'Wuhan pneumonia' and *K-pangyek* 'K-quarantine',[15] with metalinguistic information on the various components of the microstructure indicated in italics.

(14) Example of microstructure for *wuhanphyeylyem* 'Wuhan pneumonia'

wuhanphyeylyem 'Wuhan pneumonia' (武漢肺炎) compound	*Korean headword (original Hanja form) grammatical category*
Another term for COVID-19, which may be considered offensive.	*definition and register information*
The term was first used to refer to an infectious respiratory disease, which was caused by a new coronavirus identified in Wuhan, China, in December 2019, and later led to the pandemic outbreak.	*socio-cultural context*
In February 2020, the World Health Organization (WHO) announced that words indicating geographical locations should not be included in the name of	*usage information*

15 The definitions are from Nam et al. (2021: 147–148), to which additional information and examples extracted from the Comment corpus were added based on the results of the present study.



	202001	202002	202003	202004	202005	202006	202007	202008	202009	202010	202011	202012	202101	202102	202103
Article_Wuhan pneumonia	17,117	4,572	84	49	27	18	37	8	19	3	15	40	41	29	14
Comment_Wuhan pneumonia	7,584	35,797	15,357	4,207	2,583	1,721	1,141	4,057	2,713	1,786	3,717	4,545	4,486	3,848	1,575

Figure 5: Usage frequency of *Wuhanphyeylyem* ‘Wuhan pneumonia’ per genre.

the disease and officially named it 'COVID-19'	
In the case of COVID-19, it used to be called ***Wuhan Pneumonia*** after the city of Wuhan, China, where the outbreak originated.	*example from an article*
Why on earth does this disease keep changing names so often? Should we call it ***Wuhan pneumonia***? Novel coronavirus? COVID-19?	*example from a comment*
COVID-19, Wuhan corona	*synonyms*

(15) Example of microstructure for *K-pangyek* 'K-quarantine'

kheyi-pangyek 'K-quarantine' (K防疫) compound	*Korean headword (original English and Hanja forms) grammatical category*
[K: Korea]	*morphological and cultural information*
Term referring to the quarantine system implemented by the South Korean government to deal with infectious diseases and which began to be used with the global pandemic of COVID-19.	*definition*
K-quarantine has become a global standard and is rebuilding the economy faster than anything else in the world.	*example from an article*
K-quarantine budget has been drastically increased to 1.8 trillion won.	
The time of boasting about ***K-quarantine*** is long gone and now not a single vaccine for 600 infected people is a pathetic sight.	*example from a comment*
After spending 120 billion won of taxpayers' money to promote ***K-quarantine***, the medical team is worn, sickbeds are scarce and there's no vaccine prepared.	

> ***K-quarantine*** is pure propaganda for the general elections: in fact, the quarantine has failed.

The value of comment data in lexicographic description ultimately lies in the pragmatic information and the socio-cultural background it provides on headwords and which are not easily seen in existing dictionaries. Moreover, unlike articles, comments are produced by a multitude of commenters and reflect their emotions and stances in relation to the relevant neologisms, providing dictionary users and future generations with fresh, raw examples of real-life language for neologism headwords. Korean neology research has thus far focused on article texts, limiting the scope of information on neologisms. To cope with this shortcoming, it is necessary to examine the emergence of neologisms in comments and other genres so as to study and describe the various attributes of neologisms. COVID-19 neologisms, in particular, have proliferated for the past year or so, to express, describe, and comment on a global phenomenon, constituting an unprecedented case of profuse and multifaceted neological creativity centered on a single topic. This is precisely what this study sought to grasp by analyzing the differences in distribution and trends of COVID-19 neologisms across the two genres of articles and comments. Ultimately, this paper reflected on ways to apply the fruits of such research in the practical domain of lexicography and showed that the raw language of commenters, despite its many issues, has its place among language experts in dictionaries.

Bibliography

Monographs and articles

Baker, Paul (2016): The shapes of collocation. In: *International Journal of Corpus Linguistics* 21(2),139–164.

Barnhart, David K. (2007): A calculus for new words. In: *Dictionaries* 28(1),132–138.

Brezina, Vaclav/McEnery, Tony/Wattam, Stephen (2015): Collocations in context: A new perspective on collocation networks. In: *International Journal of Corpus Linguistics* 20(2),139–173.

Lee, Soojin/Kang, Hyeonah/Nam, Kilim (2020): COVID-19 Neologism Collection and Usage Pattern Analysis: A Consideration of Topic-Specific Neologism Extraction and Usage Pattern Analysis. In: *Korean Lexicography* 36, 136–171.

Metcalf, Allan (2002): *Predicting New Words: The Secrets of Their Success*. Boston: Houghton Mifflin.

Nam, Kilim (2015): An Analysis of the Usage Change of Neologisms and the Criteria for Headwords in the Dictionary – Focused on 2005–2006 neologisms. In: *Hangul* 310, 205–233.

Nam, Kilim (2019): *Neologisms of 2019*. Seoul: Nation Institute of Korean Language.

Nam, Kilim/Lee, Soojin/Choi, Jun/Seo, Eun Young/Kang, Hyunah/Baek, Mi Kyung/Jeong, Heeyun/Kim, Hae-eun/An, Jin San (2021): *Neologism 2020 – A new language of COVID-19 pandemic*. Seoul: Hankookmunhwasa.

Scott, Mike (2010): What Can Corpus Software Do? In: O'Keeffe, Anne/McCarthy,Michael (eds.): *The Routledge handbook of corpus linguistics*. New York: Routledge, 136–151.
Škrabal, Michal/Kavka, Martin (2021): Merging Professional and Collaborative Lexicography: The Case of Czech Neology. In: *International Journal of Lexicography* 34.3. ecab003 [https://doi.org/10.1093/ijl/ecab003; last access: July 30, 2021].
Stubbs, Michael (1996): *Text and Corpus Analysis: Computer Assisted Studies of Language and Culture*. Oxford: Blackwell.
Stubbs, Michael (2001): *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell publishers.
Wierzbicka, Anna (1999): *Emotions across Languages and Cultures. Diversity and universals*. Cambridge: Cambridge University Press.
Williams, Raymond (1976): *Keywords. A Vocabulary of Culture and Society*. 2015, 3rd edition. New York: Oxford University Press.

Dictionaries

Urimalsaem (2016): https://opendict.korean.go.kr/main.

Pedro J. Bueno, Judit Freixa

Lexicographic detection and representation of Spanish neologisms in the COVID-19 pandemic

1 Introduction

The syntagma *gel hidroalcohólico* 'hydroalcoholic gel' or the noun *hidroalcohol* 'hydroalcohol' cannot be found in *Diccionario de la lengua española* (DLE) of the Real Academia Española ('Royal Spanish Academy') or other general reference dictionaries of the Spanish language. This is so despite the fact that, for well over a year and to this very day, we have not been able to do anything without first sanitising our hands with this product. It is one of the many neologisms that the COVID-19 pandemic has brought us, and these have become commonly used words that dictionaries should consider as candidates for future updates.

By looking at the dictionarisability of these neologisms, in this work we try to set their boundaries on the continuum along which they fall. "Dictionarisability" means, in our context, the greater or lesser interest of these unities regarding the updating of general language dictionaries. At both ends of this continuum, there are surprising nonce words, as well as neologisms that have recently lost their status as such because they have now been incorporated into the dictionary. To identify different groups on the continuum of pandemic neologisms, we take into account the criteria proposed in the current literature and, by so doing, we are able to assess the extent to which they are discriminatory. This will allow us to address the neological process and to reflect on the various stages of it, from the time a neologism is born until the moment it ceases to be one because it has been dictionarised.

Before that, however, we present the framework of our study and refer to the mechanisms available for detecting neologisms in general and pandemic neologisms in particular.

Note: This article was prepared with the help of the LEXICAL project "Neología y diccionario: análisis para la actualización lexicográfica del español" of the Ministry of Economy and Competitiveness (ref. PID2020-118954RB-I00), funded by the State Research Agency (AEI) and the European Regional Development Fund (ERDF).

Pedro J. Bueno, Universitat Pompeu Fabra, Barcelona, Spain, email: pedrojavier.bueno@upf.edu
Judit Freixa, Universitat Pompeu Fabra, Barcelona, Spain, email: judit.freixa@upf.edu

https://doi.org/10.1515/9783110798081-005

2 Study framework

2.1 Defining what we understand by *pandemic neologism*

In most of the neology literature in Spanish, a neologism is considered to be a new word, either formally or semantically, or taken from another language.[1] However, we will use the following definition of a neologism: "A recent word that is in the process of becoming established in a language"[2] (Freixa, in press), so not all recent words are neologisms unless first signs of their common use by speakers can be noted in corpus data.

In this case, the words that we consider 'recent' are the following: a) all the terms related to the COVID-19 pandemic that first appeared between January 2020 and June 2021, and b) those that had appeared earlier but have experienced a big increase in use during the pandemic.

2.2 Detection of *pandemic neologisms*

Novelty is the main characteristic of neologisms and, since novelty is a perceptively subjective quality, a methodological criterion must be established to obtain data objectively. This criterion will necessarily be separate from the theoretical understanding of the concept of neologism. Moreover, it will always be an unsatisfactory one because we will be trying to square the circle. Assuming these limitations, the most reliable criterion for the detection of neologisms will be the comparison of analysis texts (necessarily current, these texts are the ones from which neologisms are expected to be extracted) with an exclusion corpus that must be capable of being deemed representative of the language. Ideally, this corpus should be a balanced body of texts in terms of discursive genres, themes and linguistic varieties, and it should include historical and current language. Thus, all the lexical units documented in the current texts that do not appear in the corpus deemed representative of the language may be considered new.

However, most neology observatories around the world do not have such an ideal corpus or the equipment to exploit it, so an exclusion corpus usually employs a lexicographic corpus composed of one or more dictionaries deemed representative of the language on which work is being done. In the case of Spanish neology

1 The origin of this definition can be found in the early authors of French lexicology, who faced the challenge of defining such an undefinable concept, such as Matoré (1952), Guilbert (1975) and Rey (1976).

2 Our definition is clearly inspired by Hohenhaus (2007: 18) who argues that neologisms are "words that are 'young', diachronically speaking, but which nevertheless have already entered the language as more or less institutionalised vocabulary items".

observatories, this method is used, and every unit from the analysed text not found in the exclusion corpus formed by DLE or other general reference dictionaries of the Spanish language are regarded as neologisms.

When the criterion for the detection of neologisms is determined in this way, it is called the *lexicographic criterion* (*for the detection of neologisms*). Criticism can easily be levelled at it (and it is widely criticised, indeed) because it does not discriminate neologisms from other words not found in the dictionaries for other reasons. As discussed in Bueno/Freixa (2020), by using the lexicographic criterion, what we actually get are *lexicographic neologisms*, some of which are true *neologisms* while others are *pseudoneologisms*. The following are considered as pseudoneologisms: a) morphologically regular and semantically transparent non-new words, whose meanings can be deduced from words and/or elements already found in dictionaries (this is the reason why dictionaries are reluctant to accept them); b) specialised lexical units (terms) that are already in the corresponding terminology dictionaries, whose novelty is simply the fact that they have entered general use; c) colloquialisms, non-recent units that dictionaries do not systematically include; d) old and new, general and specialised, frequent and occasional loanwords that, due to language policy criteria, dictionaries restrictively select for their lists of words; e) words bearing witness to an era and a place that are generally not likely to have a long course to run in society; f) localisms and dialectalisms that, again, dictionaries do not systematically include because of their lack of general use; g) nonce words, which appear for reasons that are more expressive than denominative, have a strong playful component and not necessarily with the object that they become part of general language; and h) variants, errors and other non-new units that are not found in dictionaries for various reasons and, by applying the lexicographic criterion, also become *pseudoneologisms*.

However, neology observatories are led by linguists who are well aware of these shortcomings and therefore filter neologisms by the type of research that is intended to be carried out. To do this, all lexicographic neologisms are accompanied by different pieces of information relating to linguistics (type of neologism, grammatical category, etc.), use (type of text, context, linguistic markers, frequency, etc.) and documents (relationship to words already documented, presence in other dictionaries, etc.).

Currently, the detection of neologisms is carried out using information technology tools designed for this purpose. In the case of the Barcelona Neology Observatory[3] the tool is called Buscaneo,[4] which was developed by the group itself in 2004 and is now used by all the Spanish neology observatories. Buscaneo scans the press and searches for all the words in the computerised dictionary. To those it cannot

3 https://www.upf.edu/web/obneo (last access: 10 June 2022).
4 http://obneo.iula.upf.edu/buscaneo/ (last access: 10 June 2022).

find, it applies filters to reject proper nouns and other uninteresting units. For the remaining ones, Buscaneo provides an interface allowing users to complete an entry form, adding data or information to fields that the program cannot automatically complete.

Buscaneo (like other automatic neology detectors), which is currently used to extract words from different types of written text (newspapers, magazines, Twitter), makes the task of detecting and recording neologisms considerably less onerous and offers a high degree of reliability. However, it has two limitations that, to date, can only be overcome by performing an additional manual extraction: first, such programs cannot detect semantic or syntactic neologisms (because, formally, they are already in the dictionary) or compound units (because the search strategy is monolexical); and second, they are not yet ready to work with oral-based texts, which are crucial to the study of lexical innovation because they are texts with a more spontaneous style.

2.3 The neological process

Beyond the discussion about which words are neological and which are not, we believe that, from a lexicographic perspective, it is more interesting to try to explain the neological process; a process that begins when a word is born and then becomes a unit that is sufficiently well-established in social use to be included in a general dictionary (although such formalisation may not occur for reasons specific to a particular dictionary), because neologisms at a more advanced stage of the neological process should be the first to be recorded in dictionaries.

This dynamic and complex vision of a neologism is based on the debate initiated by Bauer (1983), with the distinction of three moments in the establishment of a new word: the first occurrence, called a *nonce word*, followed by *institutionalisation* in use, and lastly by *lexicalisation*. That vision reached its culminating point with the work by Schmid (2008), who offered a much more comprehensive evolutionary process that split the evolution of a new word – from its first appearance to the end of its journey – into three stages, which he called *creation*, *consolidation* and *establishment*. At each stage, three processes take place simultaneously until the end of the road: firstly, at the structural level, lexicalisation occurs, which is the formal process from the creation of the word to its fixation; secondly, at the socio-pragmatic level, a neologism spreads among speakers and is potentially institutionalised; and thirdly, at the cognitive level, the concept is hypostatised, and speakers incorporate the lexicalised unit into their mental lexicon.

Based on Schmid's (2008) approach and Kerremans' (2015) review, Freixa (in press) tries to identify different neological behaviours. Of course, a *nonce word* comes first because it is the one that starts the process off. If it stops at that first

occurrence, it will remain as such and not be a neologism proper, precisely because it meets just a momentary expressive need.

Ephemeral neologisms come second. These are units that manage to acquire a certain frequency of use and, in accordance with Schmid (2008), also start the process at the cognitive and structural levels with hypostatisation of the concept and lexicalisation of the form. However, the process then stops because the neologism soon falls into disuse for some reason (but, ultimately, because the concept or form ceases to hold any interest for speakers).

If they do not stop at *nonce words* and are not characterised as being ephemeral, neologisms can follow the stabilisation process in different ways. Renouf (2013) referred to the evolution of neologisms as their *life-cycle*, based on the observation of their frequency. She identified several stages: birth, increase in frequency and occurrence, establishment, death and revival (2013: 182):

> The diachronic approach to the study of neologisms in text allows us to observe the existence of a measurable 'life-cycle' for each word. According to this metaphor, used by analogy with a human life-span, the life-cycle of a word is conceived as consisting of some or all of the following major stages: birth, or perhaps just first occurrence in text; possible increase in frequency and occurrence; productivity, creativity, settling down, assimilation and establishment in the language, obsolescence, possible death – and possible revival.

Similarly, in Freixa (in press) the histograms of a set of Spanish neologisms were studied and the following behaviours were identified: first, the *ideal neologism*, characterised by a sustained rise, which necessarily shows that the process has not concluded; second, the *logical neologism*, characterised by a rise and followed by stabilisation; and third, the *realistic neologism*, which rises, falls and then stabilises; and lastly, the *variable neologism*, which fluctuates between more or less pronounced rises and falls.

In this paper, we intend to show how much progress the different Peninsular Spanish pandemic neologisms detected by the Barcelona Neology Observatory have made in the neological process, and whether the behaviours observed in Freixa (in press) can be confirmed. We will also offer some examples of the lexicographic representation that some neologisms already dictionarised have received.

3 Corpus and methodology

The corpus neologisms that we analyse were obtained by manual and automatic extraction from oral texts (radio) and written texts (high circulation newspapers, magazines and Twitter accounts) using the lexicographic criterion mentioned above.

The corpus comprises 209 COVID-19-related neologisms that either appeared for the first time in 2020 and the first half of 2021, or had appeared earlier but experienced a striking increase over this period. The data were extracted from the

BOBNEO database,[5] but data relating to frequency were supplemented by consulting Factiva,[6] the world's biggest press database. In the corpus we observed how the frequency of some words was negligible or even non-existent till the beginning of the pandemic as in the case of *nueva normalidad* 'new normality' which numbered 910 occurrences in the year 2015, as a non-lexicalised placement, reaching 162,843 in 2020. We also noticed the extraordinary rise of *covid* and *coronavirus*, making up to more than two and a half million occurrences in just a year, and the emergence of some words exclusively related to the pandemic, such as *anticovid*, not presenting a real evolution and starting to be used in 2020 with a high frequency.

Based on these results, for the analysis we divided the neologisms into different groups, which form a continuum, by taking into account their frequency over the past twenty years (the chart shows the last three years only). We obtained the six groups in Table 1, following a progression in base 10. The table also shows the frequency results from the BOBNEO neologism database over the past thirty years to supplement the previous ones. As can be seen, the neologisms are fairly evenly distributed except in groups 4 and 5, where a greater concentration of cases occurs.

Table 1: Pandemic neologisms in frequency groups.

Group	F Factiva	F OBNEO	Number
1	0–9	1	21
2	10–99	1–4	26
3	100–999	1–9	27
4	1,000–9,999	1–20	67
5	10,000–99,999	1–50	40
6	+100,000	+50	28
			209

For the analysis, information on the horizontal axis of Factiva (age) was also taken into account, and neologisms were labelled according to whether they were first documented in 2020 or whether they already existed, in which case, their distribution was calculated over the years.

As we can see in the last row of Table 2, the neologisms that appeared in 2020 represent one third of the total, but the table shows how they are distributed according to their frequency of appearance: in the more frequent groups of neologisms, the percentage of new ones is 14–15%, whereas in the less frequent groups of neologisms, the percentage of new ones is higher than 80%. This correlation between age and frequency is quite logical.

5 http://obneo.iula.upf.edu/bobneo/index.php (last access: 10 June 2022).
6 https://global.factiva.com (last access: 10 June 2022).

Table 2: Age of pandemic neologisms by frequency group.

Group	Earlier	New	Number
1	1	20	21
2	6	20	26
3	20	7	27
4	53	14	67
5	34	6	40
6	24	4	28
	138	71	209

4 Analysis

For the analysis, we took our corpus of pandemic neologisms, organised into different groups by their frequency, and assumed that the more frequent they were, the more dictionarisable they would be. But, based on the most recent literature on updating of dictionaries (Metcalf 2002, Ishikawa 2006, O'Donovan/O'Neill 2008, Cook 2010, Adelstein/Freixa 2013, Freixa 2016, Nam et al. 2016, Freixa/Torner 2020, Klosa-Kückelhaus/Wolfer 2020, Bernal et al. 2020, among others), we also assumed that neologisms would have greater or lesser lexicographic interest depending on how long they had been in use (age), their denominative or stylistic function, their formation mechanism, and other aspects such as record of use.

To observe the extent to which trends in the units' dictionarisation and formation mechanism exist, we take into account the results shown in Table 3, where it is possible to see how the neologisms in each frequency group are distributed by the type of neologism in question. We do not, of course, intend to draw conclusions from a corpus of 200 examples and subgroups of such low numbers, but we do want to comment on the trends observed.

Little can be said about the first five types, since almost no examples were found, but Table 3 shows trends that are taken into account in the analysis, such as the concentration of neologisms formed by blending and neoclassical compounding in the groups where frequency is lower, the concentration of syntagmatic neologisms in the groups where frequency is higher, or the concentration of prefixed neologisms in the intermediate group.

In the analysis discussed below, we have put the six groups into three blocks due to the small corpus of examples. As we shall see, these three blocks have internal consistency: we can consider those in frequency groups 1–2 as *non-dictionarisable neologisms*, those found in groups 3–4 as *neologisms in the antechamber of dictionarisation* and, lastly, those in groups 5–6, where frequency is higher, as *dictionarisable neologisms*.

Table 3: Distribution by types of pandemic neologism, by frequency group.

Type	1	2	3	4	5	6	Total
abbreviation						1	1
conversion						1	1
initialism					1	1	2
borrowing from English		1			1	2	4
semantic change				3	1	1	5
compounding	1	3		4	1	1	10
blending	5	7	1	2	1		16
suffixation	2	3	3	6	3		17
neoclassical compounding	9	5	7	9	5	4	39
prefixation	2	3	9	19	7	7	47
syntagmatic compounding	2	4	7	24	20	10	67
TOTAL	**21**	**26**	**27**	**67**	**40**	**28**	**209**

4.1 Non-dictionarisable neologisms

In the main, the metaphor of war has been used to frame the discourse around the crisis caused by the COVID-19 pandemic. Most world leaders have done so, although some sectors, especially healthcare, have pointed out that this should not have been the mindset conveyed to the population. But it has been, and continues to be, because it has been observed that the general public reacts obediently to this approach (Sabucedo et al. 2020).

There are, however, different ways of dealing with a crisis, both socially and individually, and words that are heavily loaded with humour or criticism have also appeared in the vocabulary generated by the pandemic. Thus, rather than meeting a denominative need, some of the pandemic neologisms fulfil an expressive one that sometimes seeks to find the funny side of the situation to make it more bearable. These are nonce words.

In our corpus, nonce words account for almost a quarter of the total number of neologisms (47 out of 209). We found *pure nonce words* (group 1, 21 examples), i.e., those that have a really low frequency. But, by extending the concept of nonce word, we have also considered *disseminated nonce words* (group 2, 26 examples), i.e., those spread via social media, with a little higher frequency, although they are still occasional lexical events in the language.

More than half of these examples are formed by neoclassical compounding, a mechanism whose playfulness lies precisely in the seemingly serious and specialised result it yields (*teletrabajopatía* 'compulsive teleworking', *metacrisis*, *boeólogo -ga* 'boeologist'), or by blending, a word formation mechanism to which the literature has attributed a transgressive character (Hohenhaus 2007, Renner 2015, Winter-Froemel/Zirker 2015). In this case, the most recursive blending occurs with the *corona* element (*coronapincho* 'coronaspike', *coronahambre* 'coronahunger', *coronamiedo* 'coronafear'). Therefore, some authors refuse to consider them as neologisms (Gérard 2018, Klosa-Kückelhaus/Wolfer 2020, Bueno/Freixa 2020) while not seeking to take away their value; indeed, the study of these units allows us to find out about speakers' resources and dynamics in terms of linguistic creativity.

4.2 Neologisms in the antechamber of dictionarisation

The block of pandemic neologisms that falls in the central or mean frequency space is the most numerous one and comprises 27 group 3 neologisms (up to 1,000 occurrences in Factiva) and 67 group 4 ones (up to 10,000 occurrences in Factiva). These are, therefore, neologisms that have clearly begun the neological process, but, as we shall see in the analysis, have not yet completed it.

Social institutionalisation is certainly underway, but most have not been around long enough, as only a quarter of these neologisms had been documented previously. In this case, they are non-neological units in specialised use, and the novelty lies in their spread to general use: *azitromicina* 'azithromycin' has been documented since 1997 and has a total of 3,670 occurrences, of which 2,451 were observed in 2020 (in previous years, there were no more than 240 a year); in a lower frequency range, *apoyo respiratorio* 'breath support' has a total of 651 occurrences since it was first documented in 1995, of which 440 were observed in 2020 (in previous years, there were no more than 34 a year). Other units like these are *test serológico* 'serological test', *pluripatología* 'multipathology', *presintomático -ca* 'presymptomatic', etc. These units will most likely not complete institutionalisation in general use, and will return to specialised use, although this will depend on what happens with the pandemic we are still experiencing.

The abandonment of the neological process that some units have initiated will also depend on how the pandemic develops: ephemeral neologisms are units that disappear from use when they are dependent on a passing social phenomenon (be it a technological discovery, a health crisis or perhaps something related to the fashion world). *Covidiota* 'covidiot', *balconero -ra* '"balconer"', *telecolegio* 'teleschool', *coronabono* 'coronabond', *grupo burbuja* 'bubble group', to mention a few, may disappear from use before they become stable. But we must bear in mind that a characteristic feature of ephemeral neologisms is that their birth may occur more than once, i.e., a neologism that did not become institutionalised may have new

opportunities. *Coronavirus*, for example, has been sporadically documented in high circulation newspapers for more than 20 years, but it had an opportunity to become institutionalised in 2003, when the number of occurrences reached more than 1,000 due to the severe acute respiratory syndrome coronavirus (SARS-CoV) epidemic in Southeast Asia. However, the word's appearance became residual in just two years. A new attempt to become institutionalised occurred in 2015, with the Middle East respiratory syndrome coronavirus (MERS-CoV). Although its high frequency peak lasted only one year, *coronavirus* remained in use with about 100 occurrences per year until 2020, when it finally became institutionalised.

According to Schmid (2008), in the establishment of a word or what we call *the neological process*, besides institutionalisation in use, lexicalisation[7] occurs at the structural level and hypostatisation takes place at the conceptual level. Lexicalisation is a process of linguistic fixation of a new word's formal and semantic aspects, and thus it acquires a more precise meaning and a less variable form. This process, which is initiated with the first occurrences of a neologism, does not appear to have been completed in some of the examples making up the block of neologisms under analysis. For example, in Table 4, we can see that the neologism *distancia social* 'social distance' coexists alongside a diverse range of forms that show different degrees of social institutionalisation. These variants display the most defining semantic features of the concept, and together show that there has not yet been any formal fixation that, to some extent, lexicalisation entails (although the number of occurrences does inform us of the preferred variants in use).

Table 4: The neologism *distancia social* 'social distance' and its variants.

neologism	**occurrences in Factiva**
distanciamiento social 'social distancing'	114,386
distancia de seguridad 'security distance'	101,773
distancia social 'social distance'	68,134
distanciamiento físico 'physical distancing'	19,154
distancia interpersonal (interpersonal distance'	17,084
distancia física 'physical distance'	17,018
distancia sanitaria 'sanitary distance'	375

And lastly, the concepts denominated by these neologisms cannot be deemed hypostatised by the majority of speakers. When a speaker is faced with a new word, he or she analyses its morphological constituents. The more transparent and less ambiguous the morphological structure of the word is, the faster the process of understanding

7 See Lipka et al. (2004) for a review of concepts of the concepts of institutionalisation and lexicalisation.

it will be. And, depending on its level of semantic transparency, the formation of the new concept will be easier or harder. Such semantic transparency is determined by the frequency of the constituents, the number of existing lexemes with those constituents, and the semantic relationship between them. In addition, the information provided by the co-text and the context influences the development of the new concept (Schmid 2008). Some of the neologisms in this block are at an advanced stage of hypostatisation (*mascarilla higiénica* 'hygienic mask', *posconfinamiento* 'posconfinement', *antimascarillas* 'antimasks'), but others are not for a variety of reasons, such as the fact that they are highly specialised units (*gerontofobia* 'gerontophobia', *sobreinfección* 'overinfection', *dexametasona* 'dexamethasone').

We have therefore said that the neologisms in this block (frequency groups 3 and 4) are in the antechamber of dictionarisation because it is not yet time for them to enter it. The lexicographic interest that these units hold will depend on the course they take over the coming years, which in turn will depend on the evolution of the COVID-19 pandemic. Some of them, i.e., those bordering on the block of more frequent neologisms, are more institutionalised in use, are more lexicalised units, and a higher number of speakers has already hypostatised the concept, but the neological process has not yet been completed.

Those neologisms that succeed in completing this process will then face selection by a dictionary, in line with its internal criteria. In relation to DLE, Bernal et al. (2020) have noted that the internal balance of the dictionary ultimately determines the decision-making. So, for example, in the dictionary update, those neologisms forming a derivative series are good candidates. But, of course, the series cannot be unlimited: the words *infección* 'infection', *infectar* 'to infect' and *infeccioso -sa* 'infectious' are already in DLE. In pandemic use, however, the derivatives *sobreinfección* 'overinfection', *reinfección* reinfection', *reinfectar* 'to reinfect' and *reinfectado -da* 'reinfected' are recurrent and, since all of them are predictable derivatives, the dictionary may not consider them necessary (Bernal, 2021). The same applies to the *pandemia* 'pandemic' family (*postpandémico -ca* 'postpandemic', *prepandémico -ca* 'prepandemic', *antipandémico -ca* 'antipandemic' and its variants, and the *cuarentena* 'quarantine' family (*precuarentena* 'prequarantine', *postcuarentena* 'postquarantine', *semicuarentena* 'semiquarantine'), among others.

These neologisms are not usually included in general dictionaries and, at most, can be found in dictionaries of neologisms, especially, in those produced in digital format. This is the case with Antenario,[8] a dictionary of neologisms monthly updated by the neology groups in the net of Antenas Neológicas,[9] with unities from the different geolectal varieties of Spanish. In Antenario, more than 50 neologisms have

8 Antenario: https://antenario.wordpress.com (last access: 10 June 2022).
9 Antenas Neológicas: https://www.upf.edu/web/antenas (last access: 10 June 2022).

already been published under the thematic label of *Pandemia Covid-19* 'COVID-19 pandemic'. One of them is shown in Figure 1:

reinfectarse *v intr*

Año de la primera documentación: 2020

Definición	Volver a infectarse con el virus o la bacteria que produce una enfermedad después de haberse recuperado de ella.
Contextos	«Este caso muestra que es posible reinfectarse solo unos meses después de haberse curado de una primera infección», indicó en un comunicado el departamento de microbiología de la Universidad de Hong Kong (HKU). [*El Universal* (México), 24/08/2020]
	Ha advertido hoy de que han detectado que hay personas que han pasado la infección que no desarrollan la inmunidad y pueden reinfectarse. [*El Periódico* (España), 7/05/2020]
Nodos	ARG CHL ESP MEX PER
Diccionarios	Alvar1 • Alvar2 • Clave • DAMER • DEA • DNEA • DUE4 • LAR
+ info	Martes Neológico • Neologismo del mes

Figure 1: Example of pandemic neologism published in Antenario.

In Antenario, the choice made is one of building a blog-format dictionary with thematic, linguistic and pragmatic tags users can send their comments to. As seen in Figure 1, neologisms are accompanied by the usual information in the microstructure of a dictionary (lemma, grammatical category, definition and examples) and complementary information related to geolectal information as well as to the neologicity of the word (age and dictionaries in which they are already documented).

4.3 Dictionarisable neologisms

The 68 most frequent and, in principle, more dictionarisable neologisms can be found in this block. They are more dictionarisable because they are the most institutionalised ones in use and probably the most lexicalised and hypostatised ones too, because lexicalisation and hypostatisation come from use. This block, which includes 40 neologisms with a frequency between 10,000 and 99,999 occurrences and 28 neologisms with a frequency of at least 100,000 occurrences, also contains the highest percentage

of pre-existing neologisms (85.3% had already been documented prior to the pandemic). It is therefore a set of neologisms that meet two of the criteria that are often mentioned in the literature for the purpose of assessing their dictionarisation (Metcalf 2002, Ishikawa 2006, O'Donovan/O'Neill 2008, Cook 2010, Adelstein/Freixa 2013, Freixa 2016, Freixa/Torner 2020). Also mentioned in it are other criteria relating to use, which the pandemic neologisms in this group also fulfil, such as currency (they are current neologisms, although all the pandemic neologisms meet this criterion) and textual spread (they are used in texts of different types).

As for linguistic criteria, all the neologisms fulfil the criterion of correct formation and semantic necessity because, although most have predictable and compositional meaning (*semipresencial* 'semipresential', *gel hidroalcohólico* 'hydroalcoholic gel', *supercontagiador* 'superinfecter'), the speaker does not know its precise meaning. In fact, the most lexicalised syntagmatic neologisms are concentrated in this block; they are clearly denominative and, in this case, widespread in use: *crisis sanitaria* 'sanitary crisis', *presión hospitalaria* 'hospital preassure', *servicio esencial* 'essential services', among others. While general dictionaries have tended to restrict the incorporation of polylexematic units, DLE has gradually become more open to units like these, which become subentries of existing words.

The neologisms in this block also meet documentary criteria because most of them are listed in pandemic-themed dictionaries that have recently appeared, such as the *Diccionario de covid-19 (EN-ES)*[10] by the International Association of Medical Translators and Writers and Related Sciences (Tremédica). Thus, they are neologisms that have completed the neological process and, in fact, some of them have recently been incorporated into DLE, as we shall see. *Close contact* and *social bubble* are two of the pandemic neologisms already collected in the terminological dictionary published by Tremédica, as seen in Figure 2.

As can be seen, the lexicographic representation is different in this case, as the most important information for translators has been prioritised, precisely because Tremédica is an international association of medicine and related sciences translators. This way, as well as the equivalents in English, we can also consider the synonyms in both languages.

5 Already dictionarised pandemic neologisms

Fourteen of the neologisms in our corpus have already ceased to be neologisms according to the lexicographic criterion because they have recently been incorporated into DLE. These words are shown in Table 5, and reference is made to the frequency

10 https://www.tremedica.org/tremediteca/glosarios/diccionario-de-covid-19-en-es/ (last access: 10 June 2022).

close contact

contacto estrecho; **contacto cercano**

Sinonimia (en): *close-distance contact*; *high-risk contact*.

Sinonimia (es): contacto de alto riesgo.

Concepto: se refiere a estar dentro de un radio de 2 metros de la persona contagiada por tiempo prolongado y sin equipo de protección. Incluye a las personas que habitan bajo el mismo techo (→ *household contact*).

Ejemplo: *According to the Center for Disease Control and Prevention, close contact is defined as: a) Being within approximately 6 feet (2 meters) of a COVID-19 case for a prolonged period of time; close contact*

social bubble

burbuja social

Sinonimia (en): *bubble group*; *corona bubble*; *COVID bubble*; *COVID-19 bubble*; *germ bubble*; *quaranteam*.

Sinonimia (es): burbuja; grupo burbuja; grupo de convivencia estable [GCE].

Concepto: fórmula de aislamiento social intermedia entre relacionarse con todo el mundo y no hacerlo con nadie. Consiste en establecer un grupo cerrado de pequeño tamaño (por lo general, seis personas) con el que, sin ser convivientes, se mantiene contacto de manera recurrente.

Figure 2: Examples of pandemic neologisms collected by Tremédica.

group from our analysis for the purpose of seeing whether the dictionarised neologisms matched the more dictionarisable ones:

Table 5: Pandemic neologisms incorporated into RAE dictionary.

Covid	**6a**	medicalizar ‘to medicalize’	**4**
coronavírico -ca ‘coronaviral’	**4**	pandémico –ca ‘pandemic’	**5**
Coronavirus	**6**	positivo ‘positive’	**5**
cuarentenear ‘to quarantine’	**2a**	telemedicina ‘telemedicine’	**5**
desconfinamento ‘de-confinement’	**6**	teletrabajador ‘to telework’	**4**
desconfinar ‘to de-confine’	**4a**	teletrabajo ‘teleworking’	**6**
desescalada ‘de-escalation’	**6**	videollamada ‘videocall’	**6**

Indeed, most of the incorporated neologisms are in the higher frequency range (groups 5 and 6) although, as we can see, some of them are in the middle range (groups 3 and 4) and one is in the lower frequency range (groups 1 and 2). We will first focus our attention on the latter, the neologisms which, because of their frequency, were not the best candidates for updating the dictionary. The first, and most exceptional one is *cuarentenear* ‘to quarantine’, a verb that occurs just three times in BOBNEO and 91 times in FACTIVA, so it seems to be a nonce word that has spread to some extent. Given that the verb *cuarentenar* ‘to quarantine’ already exists in DLE,

the introduction of the verb ending in *-ear* might be linked to a willingness to provide better representation of non-peninsular varieties of Spanish, since *cuarentenear* 'to quarantine' has mostly been documented in Latin American countries.

The adjective *coronavírico -ca* 'coronaviral', the noun *teletrabajador -ra* 'teleworker' and the verbs *medicalizar* 'to medicalize' and *desconfinar* 'to de-confine' have, in our opinion, been rightly dictionarised for the reasons set out below. These, as Bernal et al. (2020) have already stated, are associated with DLE's internal criteria. On the one hand, they all have a relatively high frequency (more than 1,000 occurrences in 2020) and, on the other, they all complete a derivative series of other words that were already present or have been recently incorporated into the dictionary: *teletrabajador -ra* 'teleworker' (lower frequency) is consistent with the incorporation of *teletrabajo* 'teleworking' (but clearly inconsistent with the absence of the verb *teletrabajar* 'to telework'), and *coronavírico -ca* 'coronaviral' is relevant since *coronavirus* and certain derivatives thereof have also been incorporated. In some cases, the neologisms also meet the criterion of age: *teletrabajador -ra* 'teleworker' has been documented since 1995 and *medicalizar* 'to medicalize' since 1999, and the *cruciality* (Sheidlower 1995) of both is evident, since they are not products of a passing fad. All of them have a clear denominative function, had already been documented in specialised dictionaries, and refer to terms about which users may have some doubts regarding meaning or use (thus, for example, DLE gives two meanings for *medicalizar*: "dotar a algo, como un medio de transporte, de lo necesario para ofrecer asistencia médica" [to give something, such as a means of transport, what is needed to offer medical care] and "dar carácter médico a algo" [to give something a medical character]. Lastly, we should add that there is no characteristic in their formation that would render them unsuitable candidates for updating DLE.

The verb *desconfinar* 'to de-confine' deserves special attention. We would argue that its incorporation is justified in accordance with most of the criteria set out above, such as the completion of a derivative series: *confinar* 'to confine' and *confinamiento* 'confinement' were already in the dictionary, so the incorporation of reversible forms (*desconfinar* 'to de-confine' and *desconfinamiento* 'de-confinement') is as logical as the incorporation of other members of the same family having a similar frequency of use and cruciality would be, but which have nevertheless been left out: *preconfinamiento* 'preconfinement', *posconfinamiento* 'postconfinement', *reconfinamiento* 'reconfinment' and *autoconfinamiento* 'autoconfinement'. However, as already mentioned in previous paragraphs, the criterion of completion of a derivative series is limited by the criterion of formal and semantic predictability, which is used to reject units.

The other neologisms incorporated into DLE (Table 5) are in the higher frequency groups in the consulted corpora; some appeared in 2020 while others had occurrences in previous years, yet the cruciality of all of them has been evident during the pandemic. In descending order, with four million occurrences in Factiva, is *covid* (slightly more than *coronavirus*) and, with much lower frequencies but still in the highest frequency group, are *desescalada* 'de-escalation' (263,000) and *teletrabajo*

'teleworking' (156,000). The other dictionarised neologisms from the frequency group ranging from 10,000 and 99,000 occurrences are *desconfinamento* 'de-confinement' (56,000), *pandémico -ca* 'pandemic' (52,424), *videollamada* 'videocall' (36,800) and *telemedicina* 'telemedicine' (31,392).

In Figure 3 we can see three of the already collected pandemic neologisms in DLE:

coronavirus

Del ingl. *coronavirus,* de *corona* 'corona solar', por el aspecto del virus al microscopio, y este del lat. *corōna* 'corona', y *virus* 'virus', y este del lat. *virus* 'veneno', 'ponzoña'.

1. m. *Med.* Virus que produce diversas enfermedades respiratorias en los seres humanos, desde el catarro a la neumonía o la COVID.

cuarentenear Conjugar

1. intr. Pasar un período de **cuarentena** (‖ aislamiento preventivo por razones sanitarias). *Es más llevadero cuarentenear con alguien.*

2. tr. p. us. Poner algo o a alguien en **cuarentena** (‖ aislamiento preventivo por razones sanitarias). *Tendremos que cuarentenear el ganado. Las autoridades cuarentenearon el crucero.*

desconfinamiento

1. m. Levantamiento de las medidas impuestas en un confinamiento.

Figure 3: Three pandemic neologisms collected in DLE.

Figure 3 also shows how the lexicographic representation fits this kind of dictionary, in this case, a general Spanish language dictionary, also being an academic dictionary. This way, for *coronavirus*, a neologism that speakers could consider semantically unclear, the dictionary provides information about its origin and its usage in the medical area. For *cuarentenar* 'to quarantine' or *desconfinamiento* 'deconfinement', words formed following the word formation rules in Spanish, this information about origins is not provided but linguistic and usage information are.

DLE's rapid incorporation of these words is certainly positive. They meet various dictionarisation criteria and their frequency is high. However, in line with these criteria, many others may get the opportunity to be accepted into the dictionary in future updates: examples such as *gel hidroalcohólico* 'hydroalcoholic gel', *ensayo clínico* 'clinical trial', *distancia social* 'social distance' (or *distanciamiento social* 'social distancing', or variants deemed preferential, precisely pointing to usage) are units that are clearly denominative, even in the form of subentries, because of their

syntagmatic nature. Equally necessary are other words formed by compounding, such as *infectólogo -ga* 'infectologist', *sociosanitario -ria* 'sociosanitary', *semipresencial* 'semipresential'; by blending, such as *conspiranoico -ca*; or by initialism, such as *EPI* 'PPE' and *ERTE* 'furlough'. Likewise, the fact that other high frequency neologisms have been left out is understandable because they are descriptive syntagmas, such as those that have 'crisis' as their base: *crisis del covid* 'covid crisis', *crisis sanitaria* 'sanitary crisis', *crisis social* 'social crisis', or those with different families of derivatives, especially with *pre*-, *post*- and *anti*- attached to *covid*, *coronavirus*, *pandemia* 'pandemic' and other pandemic-related terms.

That said, they are neologisms that have become stable in use, and their incorporation into the dictionary will depend on the criteria that the dictionary applies to the units, not as neologisms but as language units. According to Torner (in press), a study for the lexicographic sanctioning of neology should consider this dual dimension and observe neological forms from this two-fold perspective. The dictionarisability of neology is a dual property acting on a two-fold plane: that of consolidation in use on the one hand, and that of the criteria governing the elaboration of dictionaries on the other (Torner in press).

6 Conclusions

In her magnificent work published in 2015, Kerremans compared neologisms to casting show winners: some become stable or consolidated as singers, others get to have a hit, yet most fall into oblivion. The television industry provides a context within which they can gain huge popularity within a very short space of time, but as the focus of the industry's interest shifts, the artists' popularity may quickly fall. Some manage to keep going for a while, while others manage to break into the industry without even winning the contest, so there does not appear to be a recipe for guaranteed success (Kerremans 2015: 15).

Indeed, the most dictionarisable neologisms are those with certain characteristics, yet reality has shown us time and again that many of the neologisms that fulfil those seemingly essential characteristics may not become stable, while others that do not fulfil them may.

The pandemic has mobilised vocabulary in an unprecedented way, as noted by Pons (2020) and, just 20 days into the first lockdown, words that had not previously existed began to appear, words that had not been used for a long time were revived (*lexical resurrection*, according to Pons, such as the verb *desescalar* 'de-escalate'), or a new sense or a more specific meaning was given to words already in use.

We do not know how many neologisms have been created since the start of the pandemic, but there are undoubtedly many more than the 209 analysed in this work, based on the Neology Observatory's extraction of neologisms from oral and written

texts. Such extraction has been performed annually since 1989. It provides a snapshot of how the lexicon of the language has developed to adapt to the changes in society. However, that extraction is not systematic, and although the most frequent neologisms have been detected because of their recurrent appearance in the press, many of the more fleeting ones have not. Had they been detected, the latter would have considerably enlarged our corpus of pandemic neologisms. Nonetheless, with the corpus available to us, we have been able to see that new words did appear, others were reborn, and some of the already existing ones have taken a new path.

Looking at the corpus from a lexicographic perspective, we divided this new vocabulary into three blocks. In the first, we found good examples of speakers' creativity in terms of meeting their more expressive and less denominative needs with nonce words, which performed their function yet held no lexicographic interest. In the second, we analysed a set of neologisms midway along the neological process, which could not be deemed stable in use and, therefore, were in the antechamber of dictionarisation; the path that these might ultimately take is unknown. And, in the last block, we observed those neologisms that had already completed their journey; some have already been lexicographically sanctioned, and others may be in due course.

Bibliography

Research literature

Adelstein, Andreína/Freixa, Judit (2013): Criterios para la actualización lexicográfica a partir de datos de observatorios de neología. Presentation at the International Congress *El diccionario: neología*, *lenguaje de especialidad, computación* (28–30 October 2013) in Ciudad de México.

Bauer, Laurie (1983): *English Word-formation*. Cambridge: Cambridge University Press.

Bernal, Elisenda (2021): El asedio lexicográfico a los derivados predictibles. In: *Revista Española de Lingüística* 51(1), 79–96.

Bernal, Elisenda/Freixa, Judit/Torner, Sergi (2020): Criterios para la diccionarización de neologismos: de la teoría a la práctica. In: *Signos* 53(104), 592–618.

Bueno, Pedro J/Freixa, Judit (2020): ¿Son neologismos los ocasionalismos? In: Guerrero, Gloria/ Pérez Lagos, Fernando (coord.): *Terminología, Neología y Traducción*. Madrid: Comares, 1–20.

Cook, Paul C. (2010): *Exploiting Linguistic Knowledge to Infer Properties of Neologisms*. Toronto: University of Toronto.

Freixa, Judit (in press): El proceso neológico. In: Bernal, Elisenda/Freixa, Judit/Torner, Sergi (coord.): *La neología del español: del uso al diccionario*. Madrid: Iberoamericana.

Freixa, Judit (2016): Reagrupación lexicográfica de neologismos frecuentes. In: Bernal, Elisenda/ Torner, Sergi (eds.): *Relaciones morfológicas y diccionario*. A Coruña: Universidade da Coruña, 95–112.

Freixa, Judit/Torner, Sergi (2020): On the Dictionarization of New Words in Spanish. In: *Dictionaries* 41(2), 131–153.

Gérard, Christophe (2018): Variabilité du langage et productivité lexicale. Problèmes et propositions méthodologiques. In: *Neologica* 12, 23–45.

Guilbert, Louis (1975): *La creativité lexicale*. Paris: Larousse.

Hohenhaus, Peter (2007): How to do (even more) things with nonce words (other than naming). In: Munat, Judith (ed.): *Lexical Creativity, Texts and Contexts*. Nottingham: John Benjamins, 15–38.

Ishikawa, Shin'ichiro (2006): When a word enters the dictionary: A data-based analysis of neologism. In: JACET Society of English Lexicography (ed.): *English* Lexicography *in Japan*. Bunkyo-ku: Taishukan, 39–52.

Kerremans, Daphné (2015): *Web of New Words: A Corpus-Based Study of the Conventionalization Process of English Neologisms. English Corpus Linguistics 15*. Frankfurt am Main: Peter Lang.

Klosa-Kückelhaus, Annette/Wolfer, Sascha (2020): Considerations on the acceptance of German neologisms from the 1990s. In: *International Journal of Lexicography* 33, 150–167.

Lipka, Leonhard/Handl, Susanne/Falkner, Wolfgang (2004): Lexicalization & Institutionalization. The State of the Art in 2004. In: *SKASE Journal of Theoretical Linguistics* 1, 2–19.

Matoré, Georges (1952): Le neologisme: naissance et difusión. In: *Le français moderne* 2, 87–92.

Metcalf, Allan (2002): *Predicting New Words*. Boston: Houghton Mifflin.

Nam, Kilim et al. (2016): The life and death of neologisms. On what basis shall we include neologisms in the dictionary? In: Margalitadze, Tinatin/Meladze, George (eds.): *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity*. Tbilisi: Ivane Javakhishvili Tbilisi State University, 389–393.

O'Donovan, Ruth/O'Neill, Mary (2008): A systematic approach to the selection of neologisms for Inclusion in a large monolingual dictionary. In: Bernal, Elisenda/DeCesaris, Janet (eds.): *Proceedings of the XIII EURALEX International Congress (Barcelona, 15–19 July 2008)*. Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra, 571–579.

Pons, Lola (2020): 'Covidiota', 'balconazis', 'cuarenpena' . . . los neologismos que nos ha traído la pandemia. In: *El Pais*, 08/04/2020.

Renner, Vincent (2015): Lexical blending as wordplay. In: Zirker, Angelika/Winter-Froemel, Esme (eds.): *Wordplay and Metalinguistic/Metadiscursive Reflection: Authors, Contexts, Techniques, and Meta-Reflection*. Berlin/Boston: De Gruyter, 119–133.

Renouf, Antoinette (2013): *A finer definition of neology in English: The life-cycle of a word*. In: Hasselgård, Hilde/Ebeling, Jarle/Oksfjell Ebeling, Signe (eds.): *Corpus Perspectives on Patterns of Lexis*. Amsterdam: John Benjamins, 177–208.

Rey, Alain (1976): Néologisme: un pseudo-concept? In: *Cahiers de lexicologie* 28(1),2–17.

Sablayrolles, Jean-François/Humbley, John (2021): Terminologie de la néologie. In: *Neologica* 15, 63–96.

Sabucedo, José M./Alzate, Mónica/Hur, Domenico (2020): Covid-19 y la metáfora de la guerra. In: *International Journal of Social Psychology* 35(3), 618–624.

Schmid, Hans-Jörg (2008): New words in the mind: concept-formation and entrenchment of neologisms. In: *Anglia* 126(1), 1–36.

Sheidlower, Jesse T. (1995): Principles for the inclusion of new words in college dictionaries. In: *Dictionaries* 16, 33–44.

Torner, Sergi (in press): Selección del lemario en los diccionarios de lengua y diccionarización de la neología. In: Bernal, Elisenda/Freixa, Judit/Torner, Sergi (eds.): *La neología del español: del uso al diccionario*. Madrid: Iberoamericana.

Winter-Froemel, Esme/Zirker, Angelika (2015): Wordplay and its interfaces in speaker hearer interaction: an introduction. In: Winter-Froemel, Esme (ed.): *The Dynamics of Wordplay*. Berlin/Boston: De Gruyter, 1–22.

Dictionaries

[BOBNEO] *Base de datos de los observatorios de neología*. Observatori de Neologia (Institut de Lingüística Aplicada, Universitat Pompeu Fabra). [http://obneo.iula.upf.edu/bobneo/index.php, accessed July 2021].
[DLE] *Diccionario de la lengua española*. Real Academia Española. [https://dle.rae.es, accessed July 2021].
[FACTIVA] Dow Jones. 1989–. *Factiva*. New York: Down Jones & Company. [https://global.factiva.com, accessed July 2021].
[Tremédica] *Diccionario de covid-19 (EN-ES)*. International Association of Medical Translators and Writers and Related Sciences (Trémedica). [https://www.tremedica.org/tremediteca/glosarios/diccionario-de-covid-19-en-es/, accessed July 2021].

Andreína Adelstein, Victoria de los Ángeles Boschiroli
Spanish neologisms during the COVID-19 pandemic: Changing criteria for their inclusion and representation in dictionaries

1 Introduction

The COVID-19 pandemic is a global event in a globalized society, and in many ways unprecedented. One of them is that, by August 2021, it is still an ongoing phenomenon, thus any analysis or description is provisional and/or contingent. Another feature is the immediate, urgent, and changing nature of events. Nevertheless, scientific research in different areas has had to speed up its processes in order to achieve results that have social impact; among these, those about linguistic description and lexicographic records.

The urgent need to account for this extraordinary reality as expressed in language, especially in lexical creativity, can be observed in the updates of language dictionaries in 2020 and the choice of words of the year, as well as the proliferation of an unusually large amount of individual or institutional inventories (for Spanish, for example, COVIDCIONARIO, Barale 2020, Lungevity Foundation 2021); stories in mainstream press, ephemeral publications and postings in social media where analysis and reflections are outlined with varying degrees of expertise. This was also the case in academic works describing such issues as productive resources or relationships between different languages, among others, which have multiplied since the end of 2020 and throughout 2021 (see Zholoboba 2021, Baharati 2020, Klekot 2021, Haddad/Moreno Martínez 2020, Mweri 2021, Carpintero/Tapia Kwiecien 2020).

In this context, where establishing a corpus of analysis can be a particularly difficult task – due to the seemingly unstoppable surge of new words that have been appearing in parallel to the different phases of the pandemic, scientific advances and social reactions to government health policies, and the global nature of the creative phenomenon – it is worth studying if criteria traditionally applied to include and treat neologisms in different types of dictionaries have changed in any way (see Barnhart 1985, Bernal/Freixa/Torner 2020, Cook 2010, Ishikawa 2006, Klosa-Kückelhaus/Wolfer 2020, O’Donovan/O’Neill 2008).

Andreína Adelstein, CONICET, Buenos Aires, Argentina, Universidad Nacional de General Sarmiento, Los Polvorines, Argentina, Universidad de Buenos Aires, Buenos Aires, Argentina, e-mail: aadelste@campus.ungs.edu.ar

Victoria de los Ángeles Boschiroli, Universidad Nacional de General Sarmiento, Los Polvorines, Argentina, e-mail: vboschir@campus.ungs.edu.ar

https://doi.org/10.1515/9783110798081-006

The aim of this work is to describe criteria used in the process of inclusion and treatment of neologisms in dictionaries of Spanish within the framework of pandemic instability. Our starting point will be data obtained by the Antenas Neológicas Network[1] (https://www.upf.edu/web/antenas), whose representation in three different lexicographic tools will be analyzed with the purpose of identifying problems in the methodology used to dictionarize – that is, how and what words were selected to be included in dictionaries and how they were represented in their entries – neologisms during the COVID-19 pandemic (sources and corpora of analysis, selection criteria, types of definition, among other aspects). Two of them are monolingual and COVID-19 lexical units were included as part of their updates: the *Antenario*, a dictionary of neologisms of Spanish varieties, and the *Diccionario de la Lengua Española* [DLE], a dictionary of general Spanish, published by the Real Academia Española [RAE], Spanish Royal Academy). The other is a bilingual unidirectional English-Spanish dictionary first published as a glossary, *Diccionario de COVID-19 EN-ES* [TREMEDICA], entirely made up of neological and non-neological lexical units related to the virus and the pandemic. Thus, the target lexis was either included in existing works or makes up the whole of a new tool located in a portal together with other lexicographic tools. Unlike other collections of COVID-19 vocabulary that kept cropping up as the pandemic unfolded, all three have been designed and written according to well-established lexicographic practices.

Our working hypothesis is that the need to record and define words which were recently created impacts the criteria for inclusion and treatment of neologisms in dictionaries about Spanish, including a certain degree of overlap of some features which are traditionally thought to be specific to each type of dictionary.

To this end, we will start by describing some of the most salient characteristics of the lexis of the COVID-19 pandemic in Spanish. Then, we will analyze the three lexicographic works. We will look at their headword selection procedures and how words are treated, in particular, with regard to what definition resources they deploy and how variation is recorded. Finally, we will discuss our conclusions about the peculiarities of the methodology found to be used in the inclusion and treatment of neologisms related to the pandemic.

1 The Antenas Neológicas Network, created in 2003, is one of the networks associated with the Observatori de Neologia of the Institut Universitari de Lingüística Aplicada, Pompeu Fabra University, whose purpose is to collect neology in order to describe the varieties of some Latin American countries, in addition to that of Spain. The European node is the Observatorio de Neología of the Universidad Pompeu Fabra, which registers neologisms of newspapers published in Barcelona but that have national circulation. The Latin American nodes are research teams from: Universidad Nacional de General Sarmiento (Argentina), Universidad de Concepción and Pontificia Universidad Católica de Valparaíso (Chile), Colegio de México (Mexico), Universidad Autónoma de Manizales (Colombia) and Universidad Femenina del Sagrado Corazón (Peru).

2 Neology of the COVID-19 pandemic: Main characteristics and impact on lexicography

The pandemic has impacted exponentially the neology of national languages in every field of human activity. Many of the neologisms are, in fact, internationalisms (*coronavirus*, *COVID-19*), which can be considered, to a large extent, an extreme case of what have been called *global linguistic variants* (Sayers 2014, Buchstaller 2008, apud Sayers 2014), that is, linguistic innovations that emerge simultaneously in very distant places, such as, for example, semantic neologisms like *aislamiento*, *confinamiento*, *cuarentena* (all three referring to 'lockdown') in different varieties of Spanish, or *microgota* (Spanish), *microdråpe* (Norwegian), *microgoccia* (Italian) (all of them equivalents of 'microdroplet'). From a lexicographic perspective, these global variants are likely to be included in different types of dictionaries covering the phenomenon, given the frequent use the media have made of them.

As a matter of fact, the lexis of COVID-19 in Spanish, as has been the case in other languages, includes lexical units that have a different diachronic status: words with a relatively low frequency of use that have been revitalized, which were already found in general language dictionaries (*barbijo* 'face mask') or that had not been included before the pandemic (*coronavirus*); non-neological terminological units that became frequent in everyday discourse (*carga viral* 'viral load', *oxímetro* 'pulse oximeter') and terminological neologisms that are rapidly used in the press (*supercontagiador* 'super-spreader'); denominative and/or stylistic neologisms from different fields and styles (*zoompleaños* 'Zoom birthday party', *covidiota* 'covidiot'); potential words or occasionalisms, of little (*coronabicho* 'coronabug') or no use (*coronahijo* 'coronachild').

How these words are recorded and treated lexicographically depends, of course, on the type of dictionary: language dictionaries will include items of almost all of these kinds (except for, perhaps, occasionalisms); language dictionaries of neologisms will also add those with a certain degree of diffusion; non-institutional or occasional glossaries (some of which claim to be "dictionaries" despite not following rigorous lexicographic practices) include mostly stylistic neologisms, ephemeral neologisms or occasionalisms. For example, the COVIDCIONARIO has, among others, *coronabirra* 'cocktail party during lockdown', *coronamiento* 'corona lie'; the *Diccionario Latinoamericano de la lengua española* features *coronabobo* 'coronamoron', *coronamor* 'coronalove', *coronanoico* 'corona paranoid', *covicheado* 'COVID infected'.[2] These informal records have had an unusual role in Spanish lexicography, which is discussed below.

---

2 The phenomenon described in ten Hacken/Koliopoulou (2020: 129) seems to have multiplied and is repeated throughout the globe: "New words are always marked. This is illustrated by the publication of lists and discussions of words in newspapers, which are attested in many languages".

A feature of particular relevance for lexicography is neological productivity, in terms of the productivity of neological processes (formal, semantic or loans), morphological productivity (productivity of affixes) or productivity of results (frequency of tokens). A quick look at the more than 300 neologisms recorded in 2020 by the Antenas Neológicas Network shows that the most productive processes have been syntagmatic compounds (*cuarentena intermitente* 'intermittent lockdown', *barbijo social* 'non-medical face mask'), prefixation (*postpandemia* 'post pandemic', *precuarentena* 'pre lockdown'), suffixation (*hidroalcohólico* 'alcohol gel' adj., *sanitizar* 'to sanitize'), acronymy (*covidivorcio* 'COVID divorce', *zoompleaños* 'Zoom birthday party') and loanwords (*coronacrash*, *zoomer*). However, neologisms such as *coronacrisis* and *coronabullying*, which made their way into Spanish soon after they were coined in English, may be perceived as originally Spanish acronyms rather than as loanwords. In some cases, it can be hard to decide whether a new word is a calque or an item formed in accordance with the morphological rules of Spanish. This is the case of *supercontagiador* 'super-spreader' and *microgota* 'microdroplet,' which defy easy classifications, as calques from English or derived words. On the other hand, as regards regional variation, the use of lexical variants which belong to a certain national variety by the press from a different region or country (which often happens when international news stories are translated and reproduced) tends to reinforce pan-Hispanic practices, despite (and to the detriment of) the pluricentric character of the language. Thus, depending on the country, different names have been adopted or are preferred to designate social isolation measures: *confinamiento* 'confinement' in Spain, or *aislamiento* 'isolation' and *cuarentena* 'quarantine' in Argentina and, to a lesser extent, Chile, Mexico, and Peru.

This raises the following questions: In what type of dictionary and to what extent should variants of syntagmatic compounds, such as *inmunidad comunitaria*, *inmunidad de rebaño*, *inmunidad colectiva, inmunidad de grupo* ('herd immunity'), be treated? What about those that make up a derivational paradigm, such as *barbijo social* 'non-medical mask', *barbijo quirúrgico* 'surgical mask', *barbijo casero* 'DIY mask'? How are neologisms that were created to address a phenomenon which is specific to this pandemic, but that may have a more general reference, such as *reconfinamiento* 'new lockdown' or *desconfinamiento* 'lifting of lockdown', treated?

Summing up, so far the features of Spanish neology about the pandemic that may have a bearing on the criteria for lexicographic treatment have been found to be: (i) global variants and influence of calques in speakers' perceptions (which may be perceived to be formed according to the rules of their own language, Klekot 2021); (ii) variants of the different varieties of Spanish that end up being used in others (e.g. *desconfinamiento* ('lifting of lockdown'), originally coined in Spain and later used elsewhere) (iii) a high degree of terminological banalization in different everyday fields, (iv) high degree of denominative, but also expressive, neology – linked to the ephemeral or occasional use of the word, (v) high productivity of

acronymy, especially with *corona-*, *COVID-*, *cuarent-*,[3] linked to stylistic neology, hence, occasional or ephemeral (and, as a result, unlikely to be included in general language dictionaries), as Navarro (2020) points out, (iv) changes in the way words circulate: words which have not been used much still get attention and diffusion through non-institutional lexicographic records (e.g. *covidiota* 'covidiot').

3 Theoretical framework

3.1 Neologisms and dictionaries

Neologisms are usually defined as new words; their novelty may lie in different aspects of the lexical item: morphosyntactic, such as *aplausazo* 'communal clapping'; semantic, such as *confinamiento* 'lockdown'; linked to loanwords, such as *pandemial* 'born during the pandemic'. Their neological nature may be determined through different parameters, which have been the object of many studies, especially in the Romance languages tradition (Barnhart 1985, Boulanger 1979, Cabré 2002, 2016, Cook2010, Guilbert 1975). Among these, the most widely cited criteria are the chronological (when they were first coined or recorded), psycholinguistic (speakers' perception of novelty), lexicographic (their inclusion in dictionaries) and formal instability (variation in their written or spoken renderings).

Schmid's definition (2008: 1) foregrounds an aspect of neologisms of particular interest when considering their inclusion in dictionaries, their "in-process" status, that justifies the claim that not every neological item can or should be included in dictionaries (i.e., not just in general language dictionaries, but also dictionaries of neologisms):

> Neologisms are not simply 'new words'. Rather, at least in theoretical terms, they are words which have lost their status as nonce-formations and are in the process of becoming or already have become part of the norm of the language [. . .], but are still considered new by most members of a speech community (Fisher 1998, 3; Hohenhaus 2005, 365). This of course implies that a word may be a neologism for one language user and familiar to another, and that in the absence of clues provided by the speaker signalling the newness of the word . . . hearers will be unsure whether either they are confronted with a new word or an existing word unfamiliar to them.

In connection to this, Adelstein/Boschiroli (2020: 296) discuss the paradoxical nature of neologisms as lexical units and how it affects lexicographic typology, which can be summed up as follows: (i) a neologism is not a full-fledged word, but must have the necessary conditions to become one in the future, (ii) the paradox also

3 They could even be thought of as affixes, as some authors have suggested about *-gate* or *-landia* ('-land'), which the DLE describes as a compositional element.

manifests in the fact that a neologism may be the creation of an individual speaker, but it is only through its use by a speech community that it acquires its neological status, and (iii) in the case of pluricentric languages such as Spanish, a lexical unit may cease to be a neologism in one country but still be one in others.

Furthermore, distinctions have been drawn between types of neologisms based on the extension of their social use, as well as between different types of neology: Guilbert's classical distinction between *discourse* and *language neology* (1975), Cabre's opposition between *ephemeral* and *lasting neology* (1989), the distinction between *neologism* and *occasionalism* proposed by Dressler (1993), apud Mattiello (2016: 115). These distinctions tend to suggest that only those neologisms that spread beyond the personal or occasional sphere of an individual speaker should be included in general language dictionaries.

Dictionaries of neologisms are characterized in dictionary typologies as restricted, mostly on account of chronological considerations.[4] They are language dictionaries that have a two-way relationship with general language dictionaries, which play a crucial role when determining the neological nature of a word. On the one hand, general dictionaries are used as reference points: a unit will be considered neological if it is not found in a lexicographic exclusion corpus (that is, the set of dictionaries used to corroborate whether the item is documented). On the other, once the headword list of a dictionary of neologisms has been drawn, inclusion in the general dictionary is still a central goal: the neologisms chosen for a dictionary of neologisms are likely to be included eventually in a general language dictionary. In other words, first, the general dictionary is an instrument that legitimizes the neologicity of words that will be included in the dictionary of neologisms, and secondly, dictionaries of neologisms are instruments that can be used to update general dictionaries.

Adelstein/Boschiroli (2020) identify three characteristics of dictionaries of neologisms. They are 'transition devices', since *some* of the lexical units they collect hold a special status: from a chronological point of view they are likely to be leaving their continuity stage and entering their final stage in their condition of neologism, in terms of Anula Rebollo (2010); 'remedial devices', since they include words that may not be neological from a chronological or psycholinguistic point of view, but which are neological from a lexicographic perspective; and 'documents', since they include words that may prove to be ephemeral and thus may never reach the status of institutionalized words. We will come back to these properties and review them after our analysis, to establish whether they are exclusive of dictionaries of neologisms in Spanish as regards COVID-19 vocabulary.

4 In this work we do not consider the multiple online collections recording ludic or occasional creations, most of which do not follow lexicographic criteria nor base their contents on accurate linguistic descriptions of the units.

3.2 Criteria for inclusion of neologisms in dictionaries

The process of including new words in a dictionary has usually been discussed almost exclusively in terms of the updating of general language dictionaries (see e.g. Barnhart 1985, Ishikawa 2006, O'Donovan/O'Neill 2008). Among the most cited criteria, we can identify stabilization (as opposed to the ephemeral character of neologisms), frequency of use (as opposed to hapaxes), dispersion of occurrence (as opposed to high frequency in a limited range of textual types) and, on the other hand, the witness nature of new words (Matoré 1953) and the need for naming that drives the creation of new words. Calculations to articulate criteria have also been proposed, e.g., Barnhardt (2007), Metcalf (2002), and Cook (2010).

As regards Spanish, on the premise that frequency of use is an a priori criterion for inclusion in dictionaries, Adelstein/Freixa (2013) study how neology observatories can contribute to the process of lexicographic update, concluding that a suitable proposal should take account of the different dimensions of lexis and combine formal (variants of forms previously included in dictionaries, formation rules, restrictions of the base and other elements), semantic (degrees of polysemy, polysemy production) and sociolinguistic (stability of use, extension of use, and naming needs), besides lexicographic, criteria. However, the chronological criterion is not made explicit; it is subsumed in the sociolinguist criterion of stability.

Freixa/Torner (2020) analyze dictionarization of neology in Spanish by carrying out a comparative study of data in connection to changes of frequency of neologisms throughout time and speakers' perceptions about their novelty. Adelstein/Boschiroli (2020) discuss criteria for inclusion of neologisms in neology specific dictionaries from a pluricentric, non-panhispanic perspective of Spanish.

Within Spanish lexicography, the issue of how the RAE includes new words in the DLE (often referred to as "words accepted by the RAE" by Spanish speakers at large) has been the focus of Bernal/Freixa/Torner (2020). They analyze criteria implicit in the inclusion of words in the DLE by focusing on neologisms with a high degree of frequency. Frequency of use is found to be necessary but not sufficient: other factors related to the internal coherence of the dictionary, such as completing derivative series and lexical sets (especially specialized lexis), representing geolectal variants and orienting normative use often take precedence. Words created in accordance with Spanish word formation rules are favoured over borrowings (see also Klosa-Kückelhaus/Wolfer 2020: 151). Another important factor is the inclusion of words that were created to satisfy naming needs, such as words related to new technologies or realities. Both internal coherence and naming needs seem to have been central in the 2020 update which includes words related to the COVID-19 pandemic, as will be discussed below.

4 Methodology

Our starting point in the lexicographic analysis of criteria for inclusion and microstructural treatment of neologisms is a list of 321 neological items detected and recorded during 2020 and 2021 by the Antenas Neológicas Network.[5] These data are collected exclusively from the written press of the six countries that make up the network; this may be regarded as a limitation in terms of diaphasic variation in relation to pandemic vocabulary, but on the other hand, it guarantees a certain degree of institutionalization, which is an essential aspect when considering the inclusion of new words in a general language dictionary.

The following information about the number of recorded occurrences, dispersion of use in all the countries and formation processes has been found to be relevant when analyzing criteria for inclusion in dictionaries:

- Total amount: 321
- Number of hapaxes: 96
- Number of items which were recorded in all six countries: 26
- Number of neologisms which were recorded or are being compiled in the dictionaries studied: 13 (DLE), 63 (*Antenario*), 87 (TREMEDICA).

In order to verify which items were exclusive of the pandemic – i.e., whose referents belong to the pandemic and are not revitalized forms or lexicographic neologisms from previous years – the following sources were checked: *Corpus del español NOW* [NOW] by Mark Davies (2012–2019)[6] and *Corpus del español del siglo XXI* [CORPES], updated in 2021, which is 40% press texts. This information should condition representation in the microstructure. For instance, *coronavirus* was first recorded in the

5 This network follows the same methodology and uses the same limited-access online platform to enter the relevant information about the neologisms detected from the main newspapers of the countries of the network (data about grammar, sources and type of neological formation) as the rest of the observatories and networks related to the Observatorio de Neología of the Universidad Pompeu Fabra (cf. https://www.upf.edu/web/antenas/metodologia). The results are later published in the open-access lexical database BOBNEO (http://obneo.iula.upf.edu/bobneo/index.php). A lexicographic criterion is applied for identification: the items recorded have not been included in the dictionaries that make up the exclusion corpus for each country or region, while every node checks the words against DLE and LEMA (https://www.upf.edu/web/antenas/corpus-lexicografico-de-exclusion).

6 Corpus NOW has about 7.2 billion words of data from web-based newspapers and magazines from 2012–2019.

CORPES in 2006,[7] which means it would only be neological in the SARS-Cov-2 meaning.[8] Its high frequency of use during the pandemic calls for the lexicographic inclusion of this originally specialized item in all its senses. The words documented in NOW belong to texts collected before 2019, therefore words identified as pandemic vocabulary should have a non-exclusive treatment: some examples are *aerosolización* 'aerosolization', *aerosolizar* 'to aerosolize', *aislamiento sanitario* 'sanitary isolation', *aislamiento social* 'shielding', *alcohol en gel* 'alcohol-based gel".

With the aim of determining if the users' perspective (i.e. the needs of general users) was one of the criterions when considering the inclusion of new items, we focused on the number of searches of individual items made by users of the DLE between August 2020 and August 2021, as recorded in the "Registro de consultas al diccionario de la lengua española" (/https://enclave.rae.es/herramientas/registro-de-consultas-al-diccionario-de-la-lengua-espanola-dle). These searches can also be considered an index of the degree of institutionalization of the items in the framework of the pandemic.

5 Analysis

The analysis of the lexicographic representation of COVID-19 neologisms, whose objects have changed in terms of their properties and the methodology used to study them, can be approached from two perspectives: (i) the neological processes themselves and how they are recorded and (ii) how neologisms have been dictionarized. In this section we will focus on the latter.

7 ***Coronavirus***
Absolute frequency: 1.380 Documents: 726 Normalized frequency: 4,12 cases per million)

Period	Freq	Fnorm.
2016–2020	1.300	30,99
2011–2015	38	0,46
2001–2005	34	0,33
2006–2010	8	0,07

8 These, however, are limited results: none of the following are recorded: *acuarentenamiento*, *aerosolizacion*, *aerosolizar*, *antiCOVID*, *anticuarentena*, *antipandemia*. There are 5 cases of *cuarentenar*, 2 of which are wrongly labelled as verbs, 3 of *antivacunas*, just 1 from 2020.

5.1 Bilingual unidirectional dictionary: *TREMEDICA*

The *Diccionario de COVID-19 (EN-ES)* [TREMEDICA] is an online one-way bilingual dictionary first published in May 2020 as a glossary (*Glosario de COVID-19 EN-ES)* in the webpage of TREMEDICA (https://www.tremedica.org/), an international organization that groups together translators and writers specialising in medicine and health care. Its 2.01 (June 2021) version has 6,153 headwords. The reason to publish a glossary barely two months after the pandemic had been declared was, according to the authors, to record "[not only] the spontaneous creation of many neologisms in social media, but also the widespread use of a large amount of technicisms in texts of all kinds" (Saladrigas et al. 2020: 110–111). Although bilingual, it is the largest systematic lexicographic collection of COVID-19 vocabulary in Spanish, which, as will be discussed below, has had a probably unintended impact monolingual lexicography.

TREMEDICA collects "basic terminology around COVID-19 in English" covering different aspects related to the pandemic, including lexis created and popularized in social media, to provide Spanish equivalents. This means that, on the one hand, neither the English headwords nor their translations are always neological, and on the other, part of the equivalences are proposals which, as is often the case in bilingual dictionaries, do not claim to have been attested in use. Although this – i.e. the fact that the dictionary does not necessarily reflect actual language use in Spanish – may be seen as a shortcoming from a linguistic point of view, the dictionary clearly serves and has served an extremely useful practical purpose for its intended users – translators, interpreters, journalists and other writers, especially science writers – since, given its breadth and depth of coverage, as well as the lack of other reliable lexicographic works around the subject (Navarro 2020: 790), it is a crucial reference tool that contributes to organizing and guiding lexical choices in a situation when this is highly needed. Unlike the other general monolingual language dictionaries explored, the dictionary has a functional, user-oriented focus (Tarp 2008: 47): it is mostly aimed at production and translation from English by professionals belonging to a specialized field. Therefore, although normative issues are addressed, especially through usage notes, communicative considerations seem to take precedence. This affects both macro and microstructural decisions.

5.1.1 Neologisms in the Macrostructure of *TREMEDICA*

The headword list cannot be accessed in full from the homepage, but a sample of 810 entries – which gives a good insight into the longer list – is available as an "[a] bridged glossary of COVID-19 terms (en-es)" (Saladrigas et al. 2020). This covers the lexis of "the molecular biology of coronaviruses, clinical features of COVID-19, coronavirus detection tests, diagnostic imaging tests, protective equipment, and the

COVID-19 vaccines being developed, as well as unusual neologisms, with particular emphasis on terms that are difficult to translate" (Saladrigas et al. 2020: 111). There is no explicit explanation regarding the sources where the English headwords or Spanish equivalents have been extracted from, nor any specification as to criteria for lemma selection, other than relevance to the target user. It may be assumed the source texts and corpora are listed under "Bibliografía" ('Bibliography'), though it is not clear how they are used other than the entries where examples are provided (see section 5.1.2. below). Most of the headwords and their equivalents are, in fact, either terminological (e.g. *alveolar exudate* 'exudado alveolar') or related to health care (*death toll* 'número de muertos'), not neological for the field, and unlikely to have been collected from non-specialised texts; these are clearly addressed mainly to the kind of professionals identified in the front matter; hence we will not focus on them here. However, there is a large group of headwords which were either coined (e.g. *corona bonds* 'coronabonos') or popularized (e.g. *anti-vaxxer* 'antivacunas'), during the pandemic and have relevance beyond the medical fields. Since TREMEDICA is unidirectional, strictly speaking there is no Spanish headword list; however, both English headwords and Spanish equivalents can be accessed from the same search box and are given the same label, "término" ('term'). As will be seen when the microstructure is discussed, both English and Spanish units are analyzed and explained in the entries.

Both in the English headword list and in the Spanish equivalents there seem to be few restrictions as to the type and form of neological lexical unit presented:
- in terms of length, there are single and multiword units, both in English and Spanish (see examples above)
- in terms of formation, there are words derived by affixation, composition, and acronyms (*infector* 'contagiador', *infoveillance* 'infovigilancia', *superspreader* 'supercontagiador')
- in terms of type, there are examples of different kinds of borrowings: semantic loans (*contact*, 'contacto'), calques (*plandemic* 'plandemia') and adapted loanwords (*anti-mask* 'antimascarilla') (Marello 2020: 170).

This wide range of forms and types seems consistent with a production-focused approach, characteristic of bilingual dictionaries in general and specialized dictionaries in particular, which tend to pay more attention to user communicative needs than monolingual general language dictionaries. There is one big exception, however: in accordance with RAE normative recommendations, direct loanwords used in texts in Spanish tend to be avoided, even in cases where the borrowed variant is definitely more frequent than the calque ('fake news', 'homeschooling'); so are calques which have been discouraged by RAE itself (see discussion on *sanitizar*, 'sanitize' below). However, these represent a small percentage of words associated with the pandemic. Overall, there is a conception of lexical unit that takes account of the role of multiword units in the lexis.

As pointed out earlier, equivalences are often translation solutions proposed by the dictionary, anticipating probable needs of professional users, rather than actual uses. This happens frequently when the headword is a multiword unit, where, as is common practice in bilingual lexicography, paraphrases are given, especially when no equivalents are available. This is, for example, the case of *corona-shame,* translated with the near equivalent definitional paraphrase "recriminar (una conducta que podría favorecer el contagio del coronavirus" 'to reproach (someone for a behavior that could contribute to spreading coronavirus).

In some cases, equivalents coined in accordance with word formation rules of Spanish that result in calques from English are given, although, as is the case with *corona-snitch*, these have not been found to be used in Spanish texts (see Figure 1).[9]


corona-snitch

coronachivato; **coronasoplón**; **coronabuchón**[AR]

Sinonimia (es): covichivato.

Concepto: persona que avisa a la policía para denunciar las covidioteces del coronaburro de turno (→ *covidiot*).

Ejemplo: *Corona snitches: Police create web portal to inform on neighbours* {Zindulka, 2020} <Coronachivatos: la policía ha creado un ciberportal para delatar a los vecinos>.

Tremédica-Cosnautas 2020 ©

Figure 1: *Corona-snitch* entry (TREMEDICA).

As is to be expected, given the circumstances under which the dictionary was compiled, some of the proposals were either not taken up, or not used in all geolectal varieties (when diatopic variation is not signalled, the equivalence may be assumed to work for all varieties, which is not always the case) and other emerged which seem to have become more widespread. This is the case of *fever clinic*, for example: the equivalent given is "puesto de detección (temprana) (del coronavirus)" '(coronavirus) (early) detection center", perhaps an early paraphrase solution. However, the equivalent 'unidad febril', used in Argentina, is not provided, probably because it had not been coined yet when the entry was first published.

9 "Concept: person who warns the police about the covidiocies of the latest coronamoron (see covidiot)" (our translation).

Some items which are not presented as equivalents, but as synonyms, may be regarded as exchangeable by the user, as will be discussed in the following section. Overall, it is clear that a user-oriented approach overrides other considerations and allows for the inclusion of neologisms with different degrees of institutionalization.

5.1.2 Neologisms in the microstructure of *TREMEDICA*

For every headword in English (in blue italics), TREMEDICA offers one or more equivalents in Spanish (in black bold type, see Figures 1–3); this is the only piece of information which appears in every entry in the dictionary. Often, besides the headword and the equivalent, English – "Sinonimia (en)" – and/or Spanish – "Sinonimia (es)" – synonyms are provided; these, as in the example of *essential workers*, can also work as equivalents of the lemma (see Figure 2).

essential workers

personal esencial; **trabajadores esenciales** (o trabajadoras esenciales)

Sinonimia (en): *key workers*.
Sinonimia (es): trabajadores (o trabajadoras) de actividades esenciales.
Ejemplo: *Another concern, depending on how the spread of the virus evolves, could be high rates of absentees among health care workers and other essential workers* {Washburn, 2020} <Otro motivo de preocupación, según cómo evolucione la propagación del virus, podrían ser las elevadas tasas de ausentismo de los profesionales de la salud y otros trabajadores esenciales>.

Tremédica-Cosnautas 2020 ©

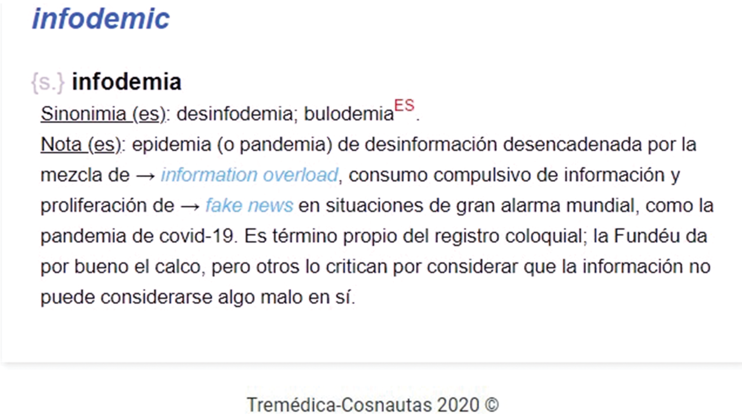
Figure 2: *Essential-workers* entry (TREMEDICA).

Thus, for every headword, an entry may suggest several equivalents, which, as suggested earlier, provides the user with different alternatives. Sometimes these are geolectal variants, as *bulodemia* (marked ES because it is only used in Spain) in *infodemic*[10] (see Figure 3):

10 "NOTE (Spanish) epidemic (or pandemic) of disinformation which results from a combination of information overload, compulsive consumption of information and proliferation of fake news in highly alarming global situations, such as the COVID-19 pandemic. It is a colloquial word: Fundeu approves of the calque, but some object to it arguing information cannot be regarded as bad in itself."

Other kinds of variants, such as stylistic variants, are often included in two other optional fields in the microstructure: “Concepto” (’Concept’, see Figure 1) or “Nota” (’Note’) (see Figure 3). In the case of neologisms, these fields are alternatively used to:

– give a definition of the word, including when it has acquired a specific meaning during the pandemic (e.g. in *de-escalation*, *cabin fever*)
– classify the word (e.g. as neologism in *plandemic*)
– explain the origin and situation in which a word is used (in *curbside*, *mask-hole*), or even the referent (in *elbow bum*p)
– explain the use of the word (in *precoronavirus*, *post coronavirus*)
– provide normative recommendations (in *herd immunity*)
– provide grammatical information (in *antivaxxer*)
– inform about synonyms (in *anti-mask*)
– cross-refer to other entries for further information (in *mask*)
– provide stylistic variants (in *corona, new normal, pandemic generation*)


infodemic

{s.} **infodemia**
Sinonimia (es): desinfodemia; bulodemia[ES].
Nota (es): epidemia (o pandemia) de desinformación desencadenada por la mezcla de → *information overload*, consumo compulsivo de información y proliferación de → *fake news* en situaciones de gran alarma mundial, como la pandemia de covid-19. Es término propio del registro coloquial; la Fundéu da por bueno el calco, pero otros lo critican por considerar que la información no puede considerarse algo malo en sí.

Tremédica-Cosnautas 2020 ©

Figure 3: *Infodemic* entry (TREMEDICA).

These kinds of explanations, often extralinguistic, may apply both to the headword and the equivalent and are particularly interesting in terms of how neologisms are represented, because they show the instability and newness of the words, and the additional difficulty involved in representing for production (rather than for comprehension): equivalents are not enough. To use them properly, extralinguistic data is necessary to make informed choices – among other things, regarding institutionalization matters such as style and degree of stability.

5.2 Monolingual dictionaries: *Antenario* and DLE

In this section we will present our analysis of two monolingual dictionaries, the *Antenario*, a restricted language dictionary, and the DLE, a general language dictionary. First, we will describe the main characteristics of each dictionary and the criteria used to select headwords. Then, we will compare how neologisms are represented in the microstructure.

5.2.1 *Antenario*

The *Antenario* is an online lexicographic dictionary of neologisms from six national varieties of Spanish; it was launched in September 2018 and has published 20 new entries every month ever since, allowing for a highly isomorphic representation vis-à-vis the dynamicity of language. By July 2021, 753 entries had been published. Both the headword list and the content of the entries are based on data about neologisms used in news media, detected and collected from 2003 by the Antenas Neológicas Network in Argentina, Chile, Colombia, Spain, Mexico, and Peru. The criterion for detection of new items is lexicographic. For a detailed account of methodology and a description of the microstructure, see Adelstein/Boschiroli (2020, 2021).

The possibility of monthly updates is a highly relevant feature for a dictionary of neologisms: whatever is published is not final and can be easily changed, which reflects the neological nature of the words. Finality is in fact often mentioned as one of the defining characteristics of a general dictionary – and also one of its main shortcomings. The online format also allows for the compilation of special issues, such as the one published at the end of 2020.

Due to the pandemic, during 2020 the Antenas Neológicas Network undertook oriented searches aimed at recording neology about COVID-19 in the member countries. In December 2020 the *Antenario* published a special issue of 49 entries of neologisms linked to the pandemic, reflecting the exceptional nature of the situation we lived throughout the year. Before the end of 2021, 18 new neologisms will be published and more COVID-19 entries are expected to be included in 2022.

Neologisms in the macrostructure of Antenario

There are two main conditions for choosing headwords for the *Antenario*: (i) as mentioned above, they all come from data collected by the Antenas Neológicas Network and (ii) as the DLE is updated every year (see section 5.2.2.), each candidate headword is checked again against the DLE to make sure they have not been included after they were documented in the Antenas bank. Criteria for selecting the headwords are total frequency of use (number of occurrences in the member nodes in the Antenas Neólogicas as recorded in BOBNEO), the witness character of the

words (*mot témoin* Matore 1953, *prominent word* Metcalf 2002) and the year they were first recorded. Thanks to the adoption of these criteria candidates are guaranteed to have a certain degree of institutionalization, which is often quite high (see in Adelstein/Boschiroli 2021 a description of how these variables have been adapted since the *Antenario* was first published).

The fact that it is updated monthly, making it possible to reconsider the criteria for the compilation of the headword lists (as well as how the items are treated in the entries), has helped, in the case of the neologisms of the pandemic, to represent in a more realistic way the gradually changing nature of the productivity of resources for lexical creation.

The data bank of Antenas Neológicas Network has recorded a large amount of what so far are considered occasionalisms (e.g. *coronaburguer* 'cororna burger', *coronapizza, cuerentenauta* 'lockdown netsurfer') and ephemeral neologisms – most of them still hapax – (*cuarentenable* 'able to be locked down', *coronabulling* 'corona bullying', *coronabus* 'bus for COVID-19 infected suspects', *covidiota* 'covidiot', *poscoronial* 'post coronavirus', adj), which were deemed as unsuitable for publication in the dictionary. However, given the dynamic nature of how entries are published, if more cases were detected, they could be included in the future. For example, there have been new records of *covidiota* or *poscoronial* used in a variety of texts, which shows their distribution has changed (7 cases were documented: none of those from 2021 are mere records of the word in inventories).

From a temporal point of view, the list of COVID-19 headwords has been growing since the first one was drawn. In December 2020 a special edition of 49 pandemic lexical items was published, based on a headword list extracted from neologisms detected between March and July 2020.[11] Then, in April 2021, a short list of candidates was selected to be included in the updates of the last months of 2021, which includes neologisms documented between August 2020 and April 2021. For 2022, the new headword list will include pandemic words recorded in 2021.

The headwords were chosen according to the following criteria. First, raw frequency: neologisms which were highly occurring neologisms were privileged (*anticuarentena* 'anti lockdown', *infectadura* 'dictatorship of the infectologists') and/or documented in most of the network's member countries (*aislamiento social* 'shielding', *home office*, *inmunidad de rebaño* 'herd immunity'). Second, geolectal or graphic variants of the first choices were added, even if the number of occurrences was low (like *aplauso sanitario* and *aplausazo* 'communal clapping', *nueva convivencia* and *nueva normalidad* 'new normal' or as *post coronavirus* and *poscoronavirus*). Third, although probably ephemeral, some frequent colloquial neologisms

11 These can be accessed here: https://antenario.wordpress.com/tag/pandemia-COVID-19/. Many neologisms that had already been compiled were deleted from the original list because they were included in the DLE's 23.4 version in November 2020.

were included because they were considered to be witness words and are not hapax: *corona*, *covidivorcio* 'covidivorce', *zoompleaños* 'Zoom birthday party'.

Neologisms that will have been published by December 2021 were chosen with different criteria. One concern was to complete either the derivational series of the headwords published in December 2020 (e.g. *prepandemia* 'pre pandemic') or semantic series (*autoaislamiento* 'self-isolation', *autocuidado* 'self-care', *autoexamen* 'self-test'). Secondly, to include synonyms or regional variants that had not been documented in other lexicographic tools (e.g. *cubrebocas* 'face mask'). Finally, to offer some of the most frequent items detected after the special edition was written and published (*coinfección* 'coinfection', *semipresencial* 'partly face-to-face', *oxímetro* 'pulse oximeter').

5.2.2 *Diccionario de la lengua española* [DLE]

The *Diccionario de la lengua española* [DLE], published by the Real Academia Española (RAE), is the monolingual general language dictionary of Spanish most widely searched by both native and non-native Spanish speakers. Its current 23rd edition (first published in 2014) is updated online once a year, around November. The main changes are the inclusion of new entries and the addition of new meanings or new information to published entries. The November 2020 update (see sample in https://dle.rae.es/docs/Novedades_DLE_23.4-Seleccion.pdf) included at least 15 changes related to the pandemic.

COVID-19 pandemic neologisms in the macrostructure of the DLE

The number of changes related to the pandemic in the DLE may look scarce when compared to TREMEDICA, Antenas, or even the words the RAE itself recorded in April 2020 as the most searched in the early months of the pandemic.[12] The changes were:

- inclusion of entries: *coronavirus*, *coronavírico* ('coronavirus' adj), *COVID-19*, *cuarentenar* / *cuarentenear* / *encuarentenar* ('to quarantine'), *desconfinamiento* ('lifting of lockdown'), *desconfinar* ('to lift a lockdown'), *desescalada* ('de-escalation'), *videochat* (n 'video chat'), *videollamada* (n 'video call'), *telemedicina* ('telemedicine')
- changes in existing entries: *barbijo* ('face mask'), *confinado -da* (adj 'locked down'), *confinamiento* ('lockdown', 'confinement'), *confinar* ('to lock down'), *cuarenteno -na* (n 'quarantine'), *mascarilla* ('face mask').[13]

12 See https://www.rae.es/noticia/las-palabras-mas-buscadas-en-el-diccionario-durante-la-cuarentena.

13 See Zoholobova (2021) for a detailed description of 2020 amends and inclusions in contrast with previous editions of the dictionary.

If we look at the new inclusions, we can identify different situations regarding their degree of neologicity:

- those that were coined and/or first documented during the pandemic: *COVID-19*
- those that existed before the pandemic but did not comply with conditions for their inclusion, such as frequency or extension (see section 3.2.): the ones derived from *cuarentena*, *confinar* and *coronavirus*)[14]
- those that had been long in use but had not been included yet: *desescalada* ('de-escalation'), *videochat* (n 'video chat'), *videollamada* (n 'video call'), *telemedicina* ('face mask')

Something the vast majority of these items have in common is being formed from Spanish bases and morphemes and well-established rules of Spanish word formation. Even in the case of calques (such as *videollamada* and *telemedicina*), they may be interpreted as formed from Spanish bases, as the etymological information provided in the entries suggests. This is consistent with the RAE's recommendations about the use of neologisms at large, as discussed in section 3.2. Regardless of frequency or extension of use, the RAE has adopted a prescriptive stance and systematically discourages or rejects the use of loanwords or even calques. A good example is *sanitizar* 'to sanitize' (included in the *Antenario* but not in TREMEDICA). It is a verb which has been widely used during the COVID-19 pandemic and found in all sorts of registers (including government and other official texts), which was first recorded in CORPES in 2012 and is discussed in RAE's *Observatorio de Palabras* ('Observatory of Words', a portal devoted to answering queries about words which cannot be found in the DLE). It is, by the RAE's own admission, one of the most frequently searched words during the pandemic (see note 10); however, its use is discouraged on puristic grounds (our translation):

> The verb *sanitizar* (from English, 'to sanitize') has diffused lately, especially in the Americas. Despite this, it is advised to avoid the use of the word and its derivations (*sanitizado, sanitizante, sanitización* . . .) and choose instead patrimonial words [i.e. derived from Vulgar Latin] such as *sanear*, *higienizar*, *limpiar* or *desinfectar*. (https://www.rae.es/observatorio-de-palabras/sanitizar)

5.3 Neologisms in the microstructure of monolingual dictionaries

The fact that the pandemic was an ongoing, unstable phenomenon when the dictionaries did their COVID-19 updates also impacts features of representation at

14 All of these are documented as in use before 2019 in the press in NOW. CORPES documents pre-pandemic cases of *coronavirus*, *cuarentenar*, *desescalada*, *videochat*, *videollamada* and *telemedicina*.

microstructural level. In the following sections we will focus on two of them: definitions and treatment of geolectal variation.

5.3.1 Definitions and extension of reference

One of the most interesting aspects regarding microstructural representation of pandemic neologisms is how the relationship between the novelty of the headword and the extension of meaning has been reflected. The fact that the items being represented lexicographically are very recently created neologisms – even if some of them may have become highly frequent – requires defining words whose referential extension cannot be totally verified yet. The fact that some of these words are revitalizations (*barbijo* 'face mask') or banalizations of non-neological technical terms (*aislamiento social* 'shielding') contributes to this discrepancy between meaning intension and extension.

Although some of the items were coined out of an apparent need to name something specific in relation to the COVID-19 pandemic, the meaning can or could have a different extension. For example, although *postcuarentena* 'post lockdown' (included in *Antenario*) refers to a period after any of the COVID-19 lockdowns in 2020, this meaning may have a more general extension. Bearing in mind the componential nature of meaning, in abstract this word does not refer exclusively to a 2020 lockdown, since it could be used in the future, or even for similar lockdowns in the past. In other words, the neologism has been coined to name a particular situation but may later be used for other referents.

Notwithstanding the obvious extensions of meaning every word can have in natural languages, we have observed the following strategies to overcome this difficulty in the monolingual dictionaries analyzed here, DLE and *Antenario*:

a) Some definitions make no reference at all to the pandemic. In general, they seem to refer to words which are not neological from a chronological point of view, as queries in corpora such as NOW or CORPES attest, or are banalized technical terms. Examples of this can be found, among others, in *alcohol en gel* 'alcohol-based gel' (and its synonyms) or *supercontagiador supercontagiadora* 'super spreader' in *Antenario* (see Figure 4).[15]

15 See *supercontagiador supercontagiadora* 'super-spreader', sense 2: "2. *Adj* Aplicado a una persona infectada, que tiene la capacidad de contagiar el virus a un gran número de personas" ('Of an infected person, being able to transmit the virus to a large number of people') and example 2 "El hospital y la iglesia suponen por sí solos el 75 por ciento de los contagios del COVI –19 en Corea del Sur, que vio multiplicados casi por 30 las infecciones desde el pasado martes, cuando dio positivo la llamada "paciente 31", una seguidora de 61 años de Shincheonji que las autoridades creen que pudo actuar como agente **supercontagiador** y transmitir la enfermedad a decenas de personas. [*El Tiempo* (Colombia), 24/02/2020]" ('Hospitals and churches amount for 75 percent of COVID-19

supercontagiador supercontagiadora ***m*** **y** ***f*** **y** ***adj***

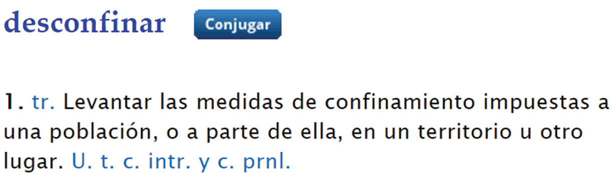
Año de la primera documentación: 2020

Definición	**1** *m* y *f* Persona infectada por un virus que tiene la capacidad de contagiar a un gran número de personas. **2** *adj* Aplicado a una persona infectada, que tiene la capacidad de contagiar el virus a un gran número de personas. **3** *adj* Aplicado a una situación o un evento, que tiene condiciones, como ausencia de ventilación o elevada concurrencia, para que se produzca un gran número de contagios de un virus.
Contextos	«Un supercontagiador es un paciente que infecta a un número desproporcionado de contactos», explica. [*Las Últimas Noticias* (Chile), 25/05/2020]
	El hospital y la iglesia suponen por sí solos el 75 por ciento de los contagios del COVI -19 en Corea del Sur, que vio multiplicados casi por 30 las infecciones desde el pasado martes, cuando dio positivo la llamada «paciente 31», una seguidora de 61 años de Shincheonji que las autoridades creen que pudo actuar como agente supercontagiador y transmitir la enfermedad a decenas de personas. [*El Tiempo* (Colombia), 24/02/2020]
	El 10 de marzo, durante el ensayo de un coro de iglesia en el estado de Washington, se contagiaron el 87% de los presentes, cuenta Lea Hammer, epidemióloga del departamento de salud pública del condado de Skagit y autora principal de un estudio que advierte sobre eventos con potencial supercontagiador, donde una persona o un pequeño número de personas deja un tendal de infectados. [*La Nación* (Argentina), 25/06/2020]
Nodos	ARG CHL COL ESP MEX PER

Figure 4: *Supercontagiador supercontagiadora* entry (*Antenario*).

infections in South Corea, whose cases have multiplied by 30 since last Tuesday, when "patient 31", a 61-year-old Shincheonji follower who is suspected to have been a superspreader agent who transmitted the diseases to dozens of people, was tested positive").

The only reference to the pandemic in the entry for *supercontagiador supercontagiadora* (Figure 4) can be found in the examples ("Contextos"). The same happens in some new entries or meanings in the DLE, as in the second sense of *confinamiento* 'lockdown', 'confinement': "2. m. Aislamiento temporal y generalmente impuesto de una población, una persona o un grupo por razones de salud o de seguridad. *El Gobierno decretó un confinamiento de un mes*." ('Temporary isolation of a community, a person or a group, often externally imposed, for health or security reasons. The Government has declared a one-month lockdown.'). A more indirect way to refer to the pandemic is including in the definition of a headword a word whose entry has an example about the pandemic. For example, the second sense in *confinado, da* 'locked down' and the new entries *desconfinar* 'to lift a lockdown' (see Figure 5),[16] *desconfinamiento* 'lifting of lockdown' all include the newly-defined word *confinamiento* .[17] However, many of the entries, amendments or additions make no reference at all to the pandemic, even when they are neologisms that are presumed to refer exclusively to the COVID-19 lockdown (*encuarentenar* 'to lock down', *COVID*).


desconfinar Conjugar

1. tr. Levantar las medidas de confinamiento impuestas a una población, o a parte de ella, en un territorio u otro lugar. U. t. c. intr. y c. prnl.

Figure 5: *Desconfinar* entry (DLE).

b) In some definitions the extension to the pandemic or other phenomena linked to it appears restricted with formulas such as "en especial . . . " or "especialmente" ('especially') or similar structures (e.g. relative clauses), since, although the words were created or revitalized during the pandemic, the reference is wider: in the *Antenario*, *aplausazo* 'communal clapping' is defined as "Acción colectiva de apoyo y reconocimiento, especialmente al personal de la salud, o de protesta, que consiste en aplaudir simultáneamente durante un período determinado" 'Colective action of support and recognition, especially of health workers', or *nueva normalidad*

---

16 *desconfinar v* 'to lift a lockdown' 1. Tr Levantar las medidas de confinamiento impuestas a una población, o parte de ella, en un territorio u otro lugar. U.t.c. intr y c. prnl. "To lift lockdown measured imposed on a community, or part of it, in a territory or any other place. Also used as intransitive and pronominal."

17 An interesting aspect of the process of synthesis used in these definitions (referring to the noun *confinamiento* and not the verb *confinar*) is that they rely on use, rather than on the base. On the other hand, the addition of senses (in *confinado*, *-da* and *confinamiento*) results in a specialization of a meaning that is somehow included in the existing first sense, which highlights both the inadequacy of the original definition, and the fact that it is a semantic neologism.

'new normal' as "Situación posterior a una crisis que implica un cambio de hábitos o expectativas en la sociedad, como la adopción permanente de medidas de prevención e higiene en el marco de la pandemia de COVID-19" 'Situation after a crisis that calls for a change in habits or expectations in society, like the permanent adoption of preventive and health measures around the COVID-19 pandemic'.

The DLE resorts to this kind of strategy indirectly only once, in the definition of *coronavirus* ("Virus que produce diversas enfermedades respiratorias en los seres humanos, desde el catarro a la neumonía o la COVID.", 'Virus that causes different respiratory diseases in human beings, from cough to pneumonia or COVID'. The reason for this may be that most of the DLE additions have a higher degree of stabilization than those in the *Antenario*, due to, on the one hand, the different nature of the dictionaries (general language vs. neologisms), and on the other, the more conservative approach to new additions the RAE favours, as discussed in 5.2.2.1.

c) Some of the items refer to events that happened during the pandemic and the definition reflects this, even when the componential meaning of the word could be used in the future for other situations or referents. For example, in *anticuarentena* 'antilockdown' all three senses refer to a reaction against "las disposiciones gubernamentales de aislamiento preventivo implementadas a causa de la pandemia de COVID-19" ('government measures of preventive isolation taken as a result of the COVID-19 pandemic') (see Figure 6). This strategy has not been found in DLE so far.

d) Unsurprisingly, specific words which are unlikely to be used with other referents or in future situations, also include references to the pandemic in their definition, e.g., *coronabono* 'coronabond' ("título de deuda común europea de emisión única creado para mitigar la crisis económica generada por la pandemia de COVID-19" 'Type of European bond . . . created to mitigate the economic crisis caused by the COVID-19 pandemic'), *poscoronavirus* 'post coronavirus' ("del periodo posterior a la pandemia provocada por el coronavirus causante de COVID-19 o relativo a él", 'of the period after the pandemic resultingfrom the coronavirus that causes COVID-19 or relative to it'). The DLE has not included this kind of headword either.

In connection to this, it is clear that the low degree of stability of the neologisms is a problem in terms of lexicographic representation since, on the one hand, they are words that can easily change meaning, in which case their definition will become outdated, and on the other, as we have seen before, their reference may change. For example, *covidivorcio* 'covidivorce' is defined in the *Antenario* as "divorcio matrimonial producido en el marco de la situación de aislamiento a causa de la pandemia de COVID-19" ('divorce that took place while in lockdown during the COVID-19 pandemic'). This definition refers indirectly to a 2020 lockdown; however, the pandemic has not finished yet and the word *covidivorcio* may end up being used to any divorce

anticuarentena *m* y *f*, *f* y *adj*

Año de la primera documentación: 2020

Definición	**1** *m* y *f* Persona que se opone a las disposiciones gubernamentales de aislamiento preventivo implementadas a causa de la pandemia de COVID-19. **2** *f* Oposición a las disposiciones gubernamentales de aislamiento preventivo implementadas a causa de la pandemia de COVID-19. **3** *adj* Que se opone a las disposiciones gubernamentales de aislamiento preventivo implementadas a causa de la pandemia de COVID-19.
Contextos	Además, el presidente de Argentina, Alberto Fernández, comparó a los anticuarentena con los terraplanistas: "Hacen mucho ruido y los medios de comunicación presentan como actores de grandes jornadas patrióticas", dijo. [*The Clinic* (Chile), 15/08/2020]
	Señalan a Patricia Bullrich y Miguel Ángel Pichetto por fogonear la anticuarentena y los acusan de estar detrás de las marchas. [*Clarín* (Argentina), 2/06/2020]
	Los primeros apuntan especialmente contra los grupos anticuarentena, que incluso protagonizaron una serie de marchas —la última y una de las más masivas el pasado 12 de octubre— protestando por la falta de libertad y lo que algunos llaman una «infectadura». [*El Universal* (México), 21/10/2020]
Nodos	ARG CHL MEX

Figure 6: *Anticuarentena* entry (*Antenario*).

in this period, and not necessarily to the ones during lockdown. This may require adjusting the definition in the future if such change were observed.

To sum up, although semantic changes are a feature of every natural language and dictionaries are regularly updated to account for them, in this case, the timing has been radically different, leading to immediacy in representation, added to the fact the events referred to in the definitions are unfinished, all of which results in problems for the lexicographic representation of neologisms, including the relative accuracy of the definitions, in other words, their decreased reliability and shorter-termed validity.

5.3.2 Geolectal variation

Another aspect of the microstructure, in the case of *Antenario*, that is affected by the unfinished nature of the pandemic is geolectal representation. Attempts were made to account for geolectal variants of COVID-19 headwords, even when not all of them were originally documented when the relevant data were collected. Also, the extension of use may have varied in different countries as the pandemic unfolded. Although in theory it would be possible to include these variations, this is difficult to do in practice given the number of changes it would involve.

As a matter of fact, the speed at which COVID-19 neologisms have been included in dictionaries affects dictionaries of neologisms – which, because of their very specificity, usually deal with phenomena which are not entirely stable – differently than other types of dictionaries. Still, the volume of new words recorded in such a short time is unprecedented.

In the case of DLE, except *barbijo*, the words and senses related to the pandemic are not marked diatopically, suggesting they are commonly used in all varieties, even if some of them were hardly used and, when they were, they were used to refer to the situation in Spain (e.g. in the Latin American nodes of the Antenas neológicas network there are no records of *desescalada*).

As regards *barbijo* 'face mask' (sense 2), diatopic labels have been updated, for example, Uruguay ("Ur"), excluded in DAMER (see Figure 7), is added (see Figure 8). As is usually the case with geolectal variants, instead of defining the word there is a

barbijo.

I. 1. m. *Bo*, *Py*, *Ar*, *Ur.* Cinta o correa que pasa por debajo de la barbilla y sirve para sujetar el gorro, el sombrero o el casco. ◆ **barbiquejo**.

2. *Bo*, *Py*, *Ar.* Pieza de tela que cubre boca y nariz, utilizada para mantener la asepsia, *generalmente por médicos y auxiliares*. ◆ **barbiquejo**.

3. *Py*, *Ar*, *Ur.* Cinta de cuero o cadena pequeña que se pasa por debajo de la quijada del caballo y que tiene sus extremos unidos a las argollas superiores del freno. rur. ◆ **barbera**.

II. 1. m. *Ar*, *Ur.* p.u. Herida en la cara.

III. 1. m. *PR.* **perdía**, ave. ◆ **barboquejo**; **berbequejo**; **berbiquejo**.

Figure 7: *Barbijo* entry (DAMER).

barbijo

De *barba.*

1. m. *Sal., Arg., Bol., Par. y Ur.* **barboquejo.**

2. m. *Arg., Bol., Par. y Ur.* **mascarilla** (|| máscara que cubre la boca y la nariz para proteger de patógenos).

3. m. *Arg.* Herida en la cara.

Figure 8: *Barbijo* entry (DLE).

mascarilla

1. f. Máscara que solo cubre el rostro desde la frente hasta el labio superior.

2. f. Máscara que cubre la boca y la nariz de su portador para protegerlo de la inhalación y evitar la exhalación de posibles agentes patógenos, tóxicos o nocivos. *Mascarilla quirúrgica, sanitaria.*

3. f. Vaciado que se saca sobre el rostro de una persona o escultura, y particularmente de un cadáver.

4. f. Capa de diversos productos cosméticos con que se cubre la cara o el cuello durante cierto tiempo, generalmente breve, con fines estéticos.

5. f. Cosmético comercial o preparado casero para regenerar, suavizar y dar brillo al cabello, que se aplica después del lavado y se deja actuar durante unos minutos antes del último aclarado. *Sería conveniente que además de un acondicionador te aplicaras una mascarilla.*

Figure 9: *Mascarilla* entry (DLE).

cross-reference to *mascarilla* ‘face mask’ (see Figure 9), but also a specification that adds information “para protegerlo de la inhalación y evitar la exhalación de posibles agentes patógenos, tóxicos o nocivos” (‘to protect from the inhalation and avoid the exhalation of possible pathogenic, toxic or harmful agents’). However, whereas in *mascarilla* multiword units headed by the noun which were frequent in everyday discourse during the pandemic are included as examples (*mascarilla quirúrgica*, *sanitaria* ‘medical mask’), no multiword units (e.g. *barbijo quirúrgico*, *barbijo social* ‘non-medical mask’) are included in *barbijo* although, as the Antenas Neológicas data show, they have been very frequent throughout the pandemic.

To sum up, in each of the lexicographic tools studied, much of the microstructural information is, to a certain extent, provisional.

6 Conclusions

In this final section we discuss the results of our analysis of how the characteristics of Spanish neology during the COVID-19 pandemic (extremely recent neologisms referring to a phenomenon still in process, which provides little time to evaluate frequency of use and degree of stabilization of the items) have impacted the criteria applied for the inclusion and treatment of neologisms in different types of lexicographic tools and, as a result, on dictionary typology and their social role.

As regards criteria for inclusion of neologisms, in the bilingual dictionary TREMEDICA, many of the items suggested as Spanish equivalents are proposals coined by the authors or are ephemeral, as documented by Antenas Neológicas. Although their inclusion may be driven by the aim to anticipate users' needs, especially translators', they are often forms which have hardly been verified in use. This can become a problem in the field of lexicography: these items are thus documented, and their documentation can be retrieved later by other lexicographic tools or the press as evidence of actual use. Furthermore, the need for urgent compilation has also impacted the lack of systematicity in the microstructure: not all entries have the same type of information in the same fields (the fields "concepto" and "nota" often seem to be used indistinctly) and the synonymous status of variants is not clear.

As regards monolingual dictionaries of Spanish, when it comes to criteria for inclusion it is apparent that relevance, dispersion of occurrence (vis-a-vis the high frequency of a narrow range of textual types), the witness nature of the items and naming needs have all been considered. However, both the chronological criterion and, more broadly, the criterion of stabilization (as opposed to the ephemeral nature of some new coinages), have not always been applied rigorously.

In the case of the DLE, questions arise about whether users' searches of what they may perceive as neologisms is a working criterion for dictionarization of a functional type. In other words, if there is interest for a certain item which is shown to be in current use, it should be included in the dictionary whereas if it is not searched, its inclusion is not justified. For example, since August 2020 no searches have been made of *acuarentenamiento* 'lockdown', *anticuarentena* 'anti lockdwon' or *antipandemia* 'anti pandemic', while there have been 4763 searches of *cuarentenar* 'to lock down'.

As for treatment in the microstructure, in TREMEDICA, the extremely new nature of the neologisms is evident in the amount of extralinguistic or usage explanations that are necessary to complete the information conventionally provided as equivalents or definitions. In our monolingual dictionaries, this is more clearly seen in the definitions. Even if the *Antenario*, as a dictionary of neologisms, includes non-fully stabilized lexical items, the resources deployed to anticipate the extension of reference of such recent neologisms are, in our view, more suitable than the ones used by the DLE.

Clearly, the degree of institutionalization of neologisms is a criterion that has been significantly influenced (one may dare say distorted) by the unfinished and unstable nature of the phenomenon of the pandemic, affecting both monolingual dictionaries analyzed for this study.

Indeed, stability and/or stabilisation seem to have been an important factor both in the selection and the definition of COVID-19 words in the DLE i.e. not just stability of form, but also the likelihood of permanence: most of the words included in the 2020 update are patrimonial words (which may be why a lower frequency word such as *encuarentenamiento* 'lockdown' is included but a widely used calque such as *sanitizar* 'to sanitize' is not) that can be used again in the future, or that could have been included in the dictionary, i.e. not restricted or tied to a transitory situation or period. The DLE thus honours the RAE tradition. However, this condition seems to be necessary but not sufficient to include words in the dictionary. DLE users' needs tend to take a back seat and prescriptive considerations are privileged.

This discussion would not be complete without including a few lines about an unexpected turn the situation took in April 2021, when the *Diccionario Histórico de la Lengua Española* [DHLE] was first published online, somehow modifying the lexicographic landscape in Spanish. In the presentation, the dictionary claims to "aim to describe every aspect (i.e. diatopic, diastractic and chronological) of the history of the lexis of Spanish" (our translation). Surprisingly, the headword list (which has been updated periodically since its first publication) includes a large number of recent lexical units, most of which are not included in the DLE and were created in 2020–21, derived from *corona-* (28) and *COVID-* (27) e.g., *coronoico* 'coronavirus negacionist', *covidilio* 'COVID affair'. Each of these are described in detail in an entry of their own, which provides, among other pieces of information, a definition, and real examples of use, as well as the number of documents the item has been found in. See, for example, the entry for *coronachivato* 'coronasnitch' (Figure 10):

Only two documents, identified as "docs. (2020–2021)" are named to support its existence and inclusion. The first one (Navarro 2020) is a light-hearted commentary about COVID-19 vocabulary by one of the authors of TREMEDICA ("The prefix corona- stands out because of its high productivity, used in more or less humorous neologisms such as *coronacrisis* [. . .] *coronachivatos* [. . .] and *coronaburrirse* 'coronabore' (practically any word, as you can see, was *coronable* in the *coronadays* of those state-of-alarm days"). The second one is *another* dictionary, TREMEDICA, which, as mentioned above, and as is common practice in bilingual lexicography, justified by user needs, often creates the equivalences, without necessarily claiming the word exists or circulates. A search on Google shows every example of use refers back to the DHLE entry, often mockingly. There is no evidence the word has been used other than in COVID-19 vocabulary inventories, not even in social media, which leaves us wondering what lexicographic methodology was used to formulate the definition in DHLE,


coronachivato, a *s.* (2020-)

coronachivato

10.ª Entrega (marzo de 2021)
Versión del 31/03/2021
Equipo Real Academia Española
Mostrando ordenación cronológica
Familia

Etim. Compuesto de *corona* y *chivato*.

Se documenta por primera vez, con la acepción 'persona que acusa o delata [a otra] por quebrantar las normas establecidas por las autoridades durante la pandemia del coronavirus', en septiembre de 2020, en un artículo de F. A. Navarro publicado en la *Revista Española de Cardiología (Madrid)*.

1. *s. m. y f.* Persona que acusa o delata [a otra] por quebrantar las normas establecidas por las autoridades durante la pandemia del coronavirus.

Sinónimo: covichivato, a

docs. (2020-2021) 2 EJEMPLOS:

2020 NAVARRO, F. A. *"Covid-19"* [23-09-2020] *Revista Española de Cardiología (Madrid)* Esp (HD)

La mayoría de los neologismos que nos trajo la pandemia covídica, no obstante, fueron de origen popular, humorísticos y pensados como flor de un día. Si en inglés los hablantes de a pie dieron en llamar al SARS-CoV-2 *the rona* o *Miss Rona* (por abreviación de *corona*), entre nosotros vimos nacer también alias coloquiales como *coronabicho*, *acojonavirus*, *cabronavirus*, *carallovirus*, *cojonavirus*, *confinavirus*, *coñazovirus* o *coronito*. Destaca por su productividad el prefijo *corona-*, usado en neologismos más o menos jocosos como *coronacrisis*, *coronabonos*, *coronacoma* (económico), *coronacompras*, *coronadivorcios*, *coronafiestas*, *coronapijos*, *coronachivatos*, *coronabilis*, *coronaplausos* y *coronaburrirse* (prácticamente cualquier palabra, como puede verse, fue *coronable* en los *coronadías* del estado de alarma).

2021 SALADRIGAS, M. V. ET ALII *"Glosario covid-19"* [01-01-2021] *Panace@: Revista de Medicina, Lenguaje y Traducción (Salamanca)* Esp (HD)

corona-snitch coronachivato; coronasoplón; coronabuchón[AR] [...] Persona que avisa a la policía para denunciar las covidioteces del coronaburro de turno [...].

Diccionario histórico de la lengua española
Real Academia Española © Todos los derechos reservados

Figure 10: *Coronachivato* entry (DHLE).

other than copying from TREMEDICA (which, in fact, offers a humorous definition, see Figure 1 and footnote 9) or basing it on formal considerations.

The hasty inclusion of such neologisms – which one may even doubt to classify as ephemeral, in many cases, since they have *never* been actually used in speech – can have the effect, as suggested above, of distorting linguistic reality. The word is assumed to exist *because* it has been included and given full treatment in a RAE dictionary (the DHLE) and many users, given media coverage, assume it has been included in the DLE.

This, in turn, and understandably, weakens credibility in the general dictionary, as was evident in comments in social media, and creates confusion, given the RAE's traditionally conservative approach (Bernal/Freixa/Torner 2020) and the fact that many other words Spanish speakers use in their everyday life are excluded (or banned) from either dictionary.

This leads us to conclude there has been circularity in Spanish lexicography, between author's neologisms and occasionalisms in connection to the COVID-19 pandemic recorded in different lexicographic tools – often resulting in the non-verification of the use of those words – the DHLE, their use in the press and their social circulation as mentions.

This is all the more striking if we consider the role dictionaries play in legitimizing language use "even though, in theory, they are only supposed to provide a description of the vocabulary used by members of a community" – particularly in the

case of historical languages such as Spanish – and as reference works that develop "the standard of a language and an identity", as pointed out in Rodríguez Barcia/ Moskowitz (2019: 3).

As ten Hacken/Koliopoulou (2020: 129) suggest, "dictionaries are used as an authority and interpreted as gatekeepers", which is why any word whose use has not been verified may still be socially regarded as sanctioned and accepted as a word belonging to the language once it is included in the dictionary.

Our claim about circularity in representation in Spanish lexicography and its impact how COVID-19 pandemic words circulated socially leads us to suggest three issues that need to be further studied: (i) marketing, (ii) the notion of neologism itself, and (iii) typology of dictionaries.

First, marketing considerations may have played a role in such circularity, modifying established criteria for inclusion (or even acknowledgment) of headwords in dictionaries such as the DHLE. As ten Hacken/Koliopoulou (2020: 129) point out: "As Kilgariff (2013: 81) notes, '[these words] might not be very important for an objective description of the language but they are loved by marketing teams and reviewers', somehow diverting the objectives of lexicography.

Second, regarding the concept of neologism itself, in the Spanish tradition the lexicographic criterion – especially vis-a-vis the DLE – plays a defining role when considering the loss of neologicity of a neological item. Inclusion in the DLE determines a word is no longer neological. This is why the *Antenario,* a tool which only deals with neologisms, ended up not publishing in their December 2020 special edition lexical items (e.g. *coronavírico -ca* 'coronavirus' adj., *COVID-19, desconfinamiento* 'lifting of lockdown',) which, from a chronological and/or psycholinguist perspective, were actually neological.

Finally, we find our starting hypothesis about the existence of a certain degree of overlap of some features which are traditionally thought to be specific to each type of dictionary, has been confirmed. Dictionaries which, unlike dictionaries of neologisms (which make no claim to finality of stability regarding the place in the language of the items collected), are not restricted to these phenomena or not supposed to collect them, ended up recording ephemeral or witness items, with a very low or null frequency of use. Those words are then defined considering an extension of reference and use that cannot be verified yet. The properties of being transition and/or remedial devices do not seem to be exclusive of dictionaries of neologisms when it comes to dealing with COVID-19 lexis.

Bibliography

Monographs and articles

Adelstein, Andreína/de los Ángeles Boschiroli, Victoria (2020): Dictionarization and lexical variation in dictionaries of Spanish neologisms. In: *Word* 66 (4),292–318. [https://doi.org/10.1080/00437956.2020.1827701].

Adelstein, Andreína/ de los Ángeles Boschiroli, Victoria (2021): Semantic Aspects of National Varieties of Spanish Dictionary of Neologisms, the *Antenario*. In: *International Journal of Lexicography* 34 (3). [https://doi.org/10.1093/ijl/ecab010].

Adelstein, Andreína/Freixa, Judit (2013): Criterios para la actualización lexicográfica a partir de datos de observatorios de neología. *Congreso Internacional El Diccionario: neología, lenguaje de especialidad, computación*. Mexico City, October, 2013. [https://repositori.upf.edu/handle/10230/34891; last access: March 10, 2019]

Anula Rebollo, Alberto (2010): Neologismos denotativos de ‘golpe’ en el español áureo. In: *FHisp* 41, 171–196.

Baharati, Adam Lal (2020): COVID 19 Neologisms in English. In: *The Saptagandaki Journal* Xi (11). 122–135.

Barnhart, David K. (1985): Prizes and Pitfalls of Computerized Searching for New Words for Dictionaries. In: *Dictionaries. Journal of the Dictionary Society of North America* 7, 253–260.

Barnhart, David K. (2007): A calculus for new words. In: *Dictionaries* 28, 132–138.

Bernal, Elisenda/Freixa, Judit/Torner, Sergi (2020): Criterios para la diccionarización de neologismos: De la teoría a la práctica. In: *Revista Signos. Estudios de lingüística* 53(104), 592–618.

Boulanger, Jean-Claude (1979): Problématique d’une méthodologie d’identification des néologismes en terminologie. In: Adda, R., et al. (eds.): *Néologie et lexicologie. Hommage à Louis Guilbert*. Paris: Librairie Larousse, 36–46.

Buchstaller, Isabelle (2008): The localization of global linguistic variants. In: *English World-Wide* 29: 15–44.

Cabré, María Teresa (1989): La neologia efímera. In: Bassot, Josep (ed.): *Miscellánia Joan Bastardas*. Barcelona: Publicacions de l’Abadia de Montserrat, 37–58.

Cabré, María Teresa (2002): La neologia avui: el naixement d’una disciplina. In: Observatori de Neologia (ed.): *Lèxic i neologia*. Barcelona: Observatori de Neologia/Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 29–41.

Cabré, María Teresa (2016): Principios y parámetros en una teoría de los neologismos. In: García Palacios, Joaquín, et al. (eds.): *La neología en las lenguas románicas. Recursos, estrategias y nuevas orientaciones*. Frankfurt: Peter Lang, 27–42.

Carpintero, Pablo/Tapia Kwiecien, Martín (2020): La lengua y el COVID-19 en los medios periodísticos digitales en Argentina: un estudio de corpus contrastivo. In: Molina Mejía, Jorge M./Valdivia Martín, Pablo/Venegas Velázquez, René A. (2020): *Actas III Congreso Internacional de Lingüística Computacional y de Corpus – CILCC 2020 y V Workshop en procesamiento Automatizado de textos y Corpus – WoPATeC 2020*. Medellín: Universidad de Antioquia, 21–23.

Cook, C. Paul (2010): *Exploiting Linguistic Knowledge to Infer Properties of Neologisms*. Ph.D Thesis, University of Toronto.

Dressler, Wolfgang U. (1993): Word-formation: Poetic licence. In: Asher, Ronald E. (ed.): *The Encyclopedia of Language and Linguistics. Vol. IX*. Oxford: Elsevier Pergamon, 5028–5029.

Freixa, Judit/Torner,Sergi (2020): Beyond Frequency: On the Dictionarization of New Words in Spanish. In: *Dictionaries: Journal of the Dictionary Society of North America* 41(1), 131–153.

Guilbert, Louis (1975): *La créativité lexicale*. Paris: Larousse.
Haddad, Amal/Moreno Martínez, Silvia (2020): COVID-19: A metaphor-based neologism and its translation into Arabic. In: *Journal of Science Communication* 19 (5)A01. [https://doi.org/10.22323/2.19050201].
Ishikawa, Shin'ichiro (2006): When a Word Enters the Dictionary: A Data-Based Analysis of Neologism. In: JACET Society of English Lexicography (ed.): *English Lexicography in Japan*, Bunkyo-ku: Taishukan, 39–52.
Kilgariff, Adam (2013): Using Corpora as Data Sources for Dictionaries. In: Jackson, Howard (ed.): The *Bloomsbury Companion to Lexicography*. London: Bloomsbury, 77–96.
Klekot, Nina (2021): Procesos de la creatividad léxica durante la pandemia de COVID-19. Un estudio contrastivo. In: *Roczniki Humanistyczne* LXIX (6), 101–114.
Klosa-Kückelhaus, Annette/Wolfer, Sascha (2020): Considerations on the Acceptance of German Neologisms from the 1990s. In: *International Journal of Lexicography* 33(2), 150–167. [DOI: 10.1093/ijl/ecz033].
Matoré, Georges (1953): *La méthode en lexicologie. Domaine français*. Paris: M. Didier.
Marello, Carla (2020): New Words and New Forms of linguistic Purism in the 21st Century: the Italian Debate. In: *International Journal of Lexicography* 33(2), 168–186 [DOI: 10.1093/ijl/ecz034].
Mattiello, Elisa (2016): Analogical Neologisms in English. In: *Italian Journal of Linguistics* 28(2): 103–142.
Metcalf, Allan (2002): *Predicting New Words. The Secrets of Their Success*. Boston/New York: Houghton Mifflin Company.
Mweri, Jefwa (2021): Corona Virus Disease (COVID-19) Effects on Language Use: An Analysis of Neologisms. In: *Linguistics and Literature Studies* 9(1), 36–47 [DOI: 10.13189/lls.2021.090105].
Navarro, Fernando A. (2020): La COVID-19 y el lenguaje medico. In: *Revista Española de Cardiología* 73(10),790–791.
O'Donovan, Ruth/O'Neill Mary (2008): A systematic approach to the selection of neologisms for inclusion in a large monolingual dictionary. In: Bernal, Elisenda/DeCesaris, Janet Ann (eds.): *Proceedings of the XIII EURALEX international congress (Barcelona, 15–19* July *2008)*. Barcelona: Institut University de Lingüística Aplicada, 571–579.
Rodríguez Barcia, Susana/Moskowitz, Andre (2019): An authentic pan-hispanic language policy? Spain as the point of reference in the Spanish Royal Academy's *Diccionario de la lengua espanola*. In: *International Journal of Lexicography* 32 (4), 498–527 [DOI: 10.1093/ijl/ecz012].
Sayers, Dave (2014): The mediated innovation model: A framework for researching media influence in language change. In: *Journal of Sociolinguistics* 18(2), 185–212.
Saladrigas, María Verónica, et al. (2020): Glosario de COVID-19 (EN-ES) abreviado. In: *Panace@* XXI (51), 110–172.
Schmid, Hans-Jörg (2008): New Words in the Mind: Concept-formation and Entrenchment of Neologisms. *Anglia* 126(1), 1–36.
Tarp, Sven (2008): *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Tübingen: Max Niemeyer Verlag.
ten Hacken, Pius/Koliopoulou, Maria (2020): Dictionaries, Neologisms, and Linguistic Purism. In: *International Journal of Lexicography* 33(2), 124–134. [https://doi.org/10.1093/ijl/ecaa011].
Zholobova, Anna (2021): Linguistic innovation during the COVID-19 pandemic: the Spanish language case. In: *XLinguae* 14(2), 331–349. [DOI: 10.18355/XL.2021.14.02.24].

Dictionaries and corpora

Antenario = *ANTENARIO. Diccionario en línea de neologismos de las variedades del español*. Ed. by Adelstein, Andreína/Freixa, Judit. [<http://www.antenario.wordpress.com>; last access: July, 30, 2021].

Barale, J., et al. (2020): Alguna *covetimologías*. In: *Ciencia Hoy* 29. 170: 10–11.

Calderón, P. (2021): *Guía de coronavirus: glosario de términos* [https://academy.doctoralia.co/blog/glosario-terminos-coronavirus, last access: July, 30, 2021].

COVIDCIONARIO = *COVIDcionario*. Ed. by Aldamiz-Echevarría, Gabriel. Electronic publication. [https://COVIDcionario.com/, last access: July, 30, 2021].

CORPES = *Corpus del Español del Siglo XXI (CORPES)*. Ed. by Real Academia Española [https://www.rae.es/banco-de-datos/corpes-xxi, last access: August, 13, 2021].

DAMER = *Diccionario de americanismos*. Ed. by Asociación de Academias de la lengua Española. Madrid: Santillana, 2010. [https://lema.rae.es/damer/, last access: July, 30, 2021].

Diccionario Latinoamericano de la Lengua Española. Ed. by Link, Daniel. UNTREF Media. [http://untref.edu.ar/diccionario/, last access: July, 30, 2021].

DLE = *Diccionario de la lengua española*. Ed. by Real Academia Española. Electronic edition, 23.4 edn. [http://www.rae.es/, last access: July, 30, 2021].

DHLE = *Diccionario histórico de la lengua española*. Ed. by Real Academia Española. Electronic publication. [https://www.rae.es/dhle/, last access: July, 30, 2021].

Lungevity Foundation. 2021. *Glosario de COVID-19*. [https://lungevity.org/for-patients-caregivers/COVID-19-and-lung-cancer/informaci%C3%B3n-sobre-coronavirus-en-espa%C3%B1ol/glosario, last access: July, 20, 2021].

NOW = *Corpus del español NOW*. Ed. by Davies, Mark. National Endowment for the Humanities, 2012–2019. [https://www.corpusdelespanol.org/now/, last access: August, 13, 2021].

TREMEDICA = *Diccionario de COVID-19 (EN-ES)*. Ed. by Saladrigas, María Verónica et al. Tremédica-Cosnautas. [https://www.tremedica.org/tremediteca/glosarios/glosario-de-COVID-19-en-es/, last access: July, 30, 2021].

Magdalena Coll, Mario Barité

Specialized voices in the 23rd edition of the *Diccionario de la lengua española*: Analysis of the COVID-19 field and its neologisms

1 Introduction

The unexpected spread of the Coronavirus has produced, among other things, linguistic and lexicographic changes with characteristics that are being studied as the pandemic unfolds. Neologisms were quickly coined and picked up around the world, while new words were created or new meanings were given to existing words.[1]

Naturally, Spanish language speakers were no strangers to this trend, which was soon examined by the media, linguistic observatories, and lexicographic works. Thus, in late 2020, when the updated edition of the *Diccionario de la lengua Española* (DLE, 23.4 2020)[2] came out, the lexicographic changes announced included several pandemic-related changes (https://dle.rae.es/contenido/actualizaci%C3%B3n-2020).[3]

1 This study was conducted in the framework of the Research Program on Terminology, Specialized Lexicography, and Organization of Knowledge proposal, financed under the call for "Research and Development Groups" of the Sectorial Commission for Scientific Research (CSIC), Universidad de la República, Uruguay (2018–2022). The researchers jointly responsible for the program are Mario Barité and Magdalena Coll.

2 The dimension of the trend has been such that, shortly after the start of the pandemic, various "Coronadictionaries" drawn up by readers, journalists, and others began circulating on social media and traditional media (see, for example, https://www.lanacion.com.ar/sociedad/coronavirus-zoompleanos-tapabocas-palabras-nacieron-o-se-nid2370269/).

3 This is an academic lexicographic work that is an essential authoritative reference. It originated as one of the leading objectives of the Spanish Royal Academy (Real Academia Española), in the framework of its foundation, in Madrid in 1713. Twenty-three editions had been published as of the year 2014. Starting with the 21st edition (1992), there was an increase in the number of meanings specific to individual Spanish-speaking countries, whose language academies are part of the Association of Academies of the Spanish Language (Asociación de Academias de la Lengua Española, ASALE), formed in 1951. The DLE's purpose is to compile the general lexicon used in Spain and in Hispanic countries throughout the world and it is aimed primarily at speakers whose mother tongue is Spanish. It is a normative work and receives more than 90 million queries each month on its online version (dle.rae.es).

Magdalena Coll, Universidad de la República, Montevideo, Uruguay,
e-mail: collmagdalena@gmail.com
Mario Barité, Universidad de la República, Montevideo, Uruguay,
e-mail: mario.barite@fic.edu.uy

https://doi.org/10.1515/9783110798081-007

Not only were new words connected with the coronavirus included, some old definitions were also revised to adapt them to the new global situation.

This is, without a doubt, an unprecedented scenario in terms of updating practices of the academic lexicography; the new technologies make it possible for the dictionary to be updated at a pace never before seen in the history of the DLE. With the pandemic, exceptional decisions have been made, considering the short time that elapsed, in many cases, between the emergence of a COVID-related word and its publication in the dictionary (cf. Battaner 2021).

There is also another factor that has changed the academic lexicographic landscape: the boost that the historical dictionary *Diccionario histórico de la lengua española* (DHLE, 2013–)[4] has received in recent years, as will be discussed below.

Academic lexicography is therefore undergoing a unique moment, both because of the pace at which the DLE is updated and because of the coexistence of that process with a renewed and dynamic DHLE. This is the situation in which lexicography has addressed and is addressing, at the academic level, the vocabulary of a pandemic of extraordinary dimensions.

Along these lines, this article has a double objective. First, it seeks to offer an initial approach, with critical notes, to the group of pandemic-related neologisms incorporated into the DLE in the year 2020. To that end, the trends in the academic dictionary's incorporation of neologisms will be reviewed, focusing in particular on specialized language neologisms. Second, the article presents the design of a research study that allows for the examination of any new words beginning with CORONA- added to the DLE and the DHLE. An assessment will be made of the particularities of the DLE and the DHLE regarding the incorporation of the new words, as well as the degree of correspondence or complementarity between the two works in this sense. This will show the complementary roles that the DLE and the DHLE are currently acquiring. In this sense, the new additions open up a debate on the treatment of neologisms in academic lexicography, in a particularly unique scenario.

This paper will thus give a brief overview of the policy for incorporating neologisms into the academic dictionary (section 2), with special attention to technical neologisms. The general characteristics of the updating practices of academic dictionaries

4 The purpose of the DHLE, formerly known as the *New Historical Dictionary of Spanish* (*Nuevo diccionario histórico del español*, NDHE), is to present in an organized manner the evolution of the Spanish lexicon over time and up to the present. It is a "complete dictionary" accessed for free on the Internet that seeks to compile the entire lexicon, covering every period and every region where Spanish is and has been spoken. In doing so, it shows the changes that words have experienced in meaning over time and even the accidental linguistic uses of a given period. It is a long-standing project that had frustrated attempts throughout the years, but whose development has received a decisive boost with the creation of the Pan-Hispanic Network of Academies, Universities, and Research Centers for the Production of the Historical Dictionary of the Spanish Language (*Red Panhispánica de Academias, Universidades y Centros de Investigación para la Elaboración del Diccionario histórico de la lengua española*) in 2021 (https://www.rae.es/dhle/).

will be addressed, with respect, in particular, to terms that emerged in the pandemic (section 3), with subsection 3.1 dealing with the general aspects of the updating process for the 23rd edition of the DLE, and subsection 3.2 with those of the DHLE. The specific research on words beginning with CORONA- is discussed in section 4, which is divided into a description of the design of the research study (subsection 4.1) and its findings (subsection 4.2). The paper concludes with some final considerations in section 5.

2 Neologisms and technical neologisms in academic dictionaries

As is well known, the definition of neologisms is somewhat controversial and its lexicographic treatment even more so. Lexicographers have, in fact, been discussing the criteria for the inclusion of neologisms in dictionaries since the field was first developed, but in recent years there has also been a specific line of theoretical research on the subject (e.g. Bernal et al. 2020: 593).

As early as 1992, Alvar provided an overview of the treatment given to neologisms in the *Diccionario de la Real Academia Española* (DRAE).[5] He argued that the general dictionary cannot incorporate all the words that emerge, "as [to be incorporated, words] require widespread use among speakers, authorities that use them, and a stability to ensure they are not birds of passage. The process may seem slow, but it is the only way." (Alvar 1992). This classic conception has clearly changed, primarily with respect to a word's stability, which is something that cannot be measured given the very short time spans separating the start of the pandemic and the DLE's incorporation of pandemic-related words.[6]

5 Historically, the main academic dictionary was known as DRAE (*Diccionario de la Real Academia Española*) but "since its last edition in 2014, the acronym DLE (*Diccionario de la lengua española*) is being furthered because of its identification and recognition in the lexicographic landscape, as this acronym is optimal and corresponds to the official name the dictionary has always had" (Moreno Moreno 2019: 86). For this reason, we use the acronyms DRAE or DLE, as appropriate.

6 Usage, a criterion considered valid for incorporating a word into the dictionary, has also undergone several re-conceptualizations since the first dictionary of the Real Academia Española (1726–1739), when it was based on the use by authors who "have treated the Spanish language with the greatest accuracy and elegance" (https://www.rae.es/obras-academicas/diccionarios/diccionario-de-autoridades-0). Modern-day usage is documented with data from a corpus that covers different registers, styles, and geographical areas. But in recent years another tool has emerged that is strongly linked to usage: the possibility of retrieving the queries made by users in the DLE regardless of whether the words searched for are in the dictionary or not. Records are now available of word searches made by users in the DLE for words not yet included in the dictionary, but for which search frequency is very high. These can be analyzed and systematized by lexicographers, and that, in turn, can influence decision-making when updating the dictionary.

As observed by Adelstein and Freixa (2013), the incorporation of a neological form into a dictionary usually responds to a combination of different criteria, including frequency, formal, semantic, and documentation criteria. Thus, the words most likely to be included in a dictionary are those that are used very frequently, highly necessary for naming purposes, internationalized, easily adaptable, with a derivative family, etc. (Bernal et al. 2020: 606). These criteria must also be pondered according to the type of dictionary in question (Bernal et al. 2020: 594).

Bernal et al. analyze the words added to the 23rd edition of the DLE, updated in 2019, to "infer the non-explicit criteria used to perform the selection and contrast them with the criteria proposed by specialized literature" (2020: 608). This paper will only consider the pandemic-related words that were added in 2020, that is, in the 23.4 edition of the DLE.

In their analysis, Bernal et al. (2020: 608) suggest that the above criteria have been taken into account unevenly in the academic dictionary. The frequency criterion does not appear to have been used as an exclusion filter, while the formal criterion has, given that all "the derived and compound words [incorporated into the academic dictionary] are correctly formed words (in the sense that they follow the rules for correctly forming words)" (Bernal et al. 2020: 609).Thus, the inclusion of word families "lends consistency to the dictionary in the sense that complete derivative series are provided or else derivative series that already had a representative in the dictionary are completed" (Bernal et al. 2020: 609). As for semantic criteria, the selection of entries added to the 2019 edition of the DLE would appear to underscore the denominative need. Bernal et al. observe that a significant number of neologisms that have an entry in the dictionary are words that belong to scientific subject areas, such as medicine, biochemistry, or architecture, which lost their specialized terminological value once they were incorporated into the general language (2020: 610). Moreover, documentary criteria do not appear to have been decisive in the inclusion of certain words in the DLE edition studied by Bernal et al. (2020). As will be seen in the next section, the criteria found by Bernal et al. for the 2019 incorporations are the same that were applied for the 2020 incorporations.

The authors conclude that the criteria that appear to have had more weight in the decisions to incorporate words are the criteria connected with the work's internal logic, an aspect that is maintained in the updated version considered here:

> [O]n the one hand, it is observed that many of the new words added complete derivative series; on the other, the lexicon of subject areas present in the dictionary is enhanced, thus completing the coverage thereof, despite the fact that in some cases they are highly specialized words, a characteristic that apparently should be pondered as a negative aspect. (Bernal et al. 2020: 613)

The treatment of specialized words in the academic dictionary merits some reflections. A brief examination of that academic work reveals that many specialized words are not labeled as such. For example, the DLE (2014) assigns the diatechnical

label '*Zool.*' to families, phyla, or groups, as in the case of echinoderms, but not to "starfish," one of its species.[7]

In this context, it is important to bear in mind the information that is provided in the forewords and introductions of the last editions of the DLE. Both the 2001 and the 2014 editions omit any reference to the nature of the diatechnical labels and their treatment. Only one (isolated and indirect) consideration is made, and that is when establishing the precedence of the labeled meanings. Neither edition provides an explicit explanation of what a diatechnical label is, and no definition of it is given in the body of the dictionary either (Barité/Blanco 2014).

With respect to the application of diatechnical labels, in the 2001 edition there is an interesting explanation of the so-called "technical words": "The dictionary includes any words and meanings from different fields of knowledge and professional activities whose current use – technical archaisms are also excluded – has gone beyond their original scope and spread to frequent or occasional use in everyday or cultivated language" (DRAE 2001: xlviii). In its 2014 edition, there is no mention of how diatechnical labels are applied. There is only an example in the sample entries in which the diatechnical label is explained as the specialized field to which the word or sense corresponds.

According to Barité/Blanco (2014), three different aspects of the treatment of specialized words in the academic dictionary can be inferred from the information in the 2001 edition of DRAE: (a) technical archaisms are excluded from the dictionary; (b) also excluded are specialized words or meanings that have not gone beyond their field and are therefore only known and used in internal communications in that field; and (c) specialized words and meanings whose use, whether frequent or occasional, has spread beyond its field are included. However, it is not always clear which words or meanings included merit a diatechnical label.

With the information available, it would appear that no other criteria were applied for the updates to the 23rd edition of the DLE, so that it can be concluded that the above criterion still applies. It is actually a very logical criterion, which accounts for the de-terminologization of certain expressions or an unusual socialization of certain words, without losing their condition of specialized word. It is reasonable to assume that in the innovative 24th edition, on which the Spanish language academies have been working for some years now, this type of information will be made explicit.

The inclusion of technical neologisms in dictionaries has a tradition of its own, which is strongly linked to the diversification of professional activities and fields of knowledge that give rise to specialized languages. There are also different real-life

---

7 A diatechnical label, also known as a "subject label", "specialization label", or "thematic label", qualifies the definition of a lexical unit by indicating the scientific discipline, technical field, profession, or specialized area it belongs to (Martínez de Sousa 2009: 147).

events, such as the pandemic that began in 2019, that force established disciplines to reallocate resources and quickly direct research toward problem-solving, in this case problems concerning the classification of a virus and the diagnosis, prevention, and treatment of the disease it causes. This type of situation is fertile ground for the rapid emergence of new terms that are quickly picked up and become widespread.

All these circumstances naturally occur in the oral and written language used by specialists among themselves. These terms, as such, may be included in the dictionary, and, if they are, they may be accompanied by the corresponding technical label and an appropriate definition. As the pandemic progresses, these terms will be used by specialists in their communications with lay persons, or between health authorities or specialized reporters and ordinary citizens. These terms will also begin to be used in everyday language, making their lexicographic treatment different: they will not necessarily carry a label indicating a specialized field and their definition will tend to be less technical.

Things are not, however, that simple or linear. Strictly speaking, specialized languages would only include scientific terms used by specialists in their specialized communications or in dissemination activities, and which, for example, are part of the formal (scientific or technical) classifications of their specialized fields. Within this framework, and from a lexicographic perspective, in the general dictionaries of a language, only those words or meanings that contain a diatechnical label could be considered specialized. Applying instead a more open criterion, the set of specialized words or meanings could be expanded to include those that directly or indirectly refer to a specialization, and that are usually jargon or technical slang words, or just constructions coined for communication among non-specialists, even when they are adopted because of their picturesque quality.

3 Academic dictionary updates in the context of the pandemic

3.1 General characteristics of the updating process of the 23rd edition of the DLE

The most recent edition of the DLE is the 23rd, and while it came out in 2014, it has been regularly updated in its online searchable version, which became available in 2015. Successive batches of modifications approved by the language academies, which will all be ultimately included in the 24th edition, have been periodically

incorporated.[8] The last such update, which is what interests us here, was announced in December 2020 and it is known as the 23.4 edition of the DLE.[9]

Like all dictionaries of such magnitude, DLE has been updated periodically since its first edition in 1780. In the eighteenth century, there were two updates; in the nineteenth, ten; and in the twentieth, eight. In the twenty-first century, the 22nd edition was released in 2001 and the 23rd edition came out in 2014, as noted above. In this sense, as of the twenty-first century, technological developments enabled a change in the very concept of updating, and, following the 23rd edition, updates are performed continuously. This has generated different and renewed versions of the edition published originally on paper in 2014. The dictionary is no longer something static; rather, changes can be made regularly to it and released as they are made. This is an unprecedented situation for Spanish academic lexicography.

The words connected with the pandemic that were added to the DLE as of December 2020 can be classified according to different aspects (cf. Battaner 2021): they can be considered hereditary words already present in the DLE (such as *distanciamiento* 'distancing', *normalidad* 'normality', *barbijo* 'mask', *aislamiento* 'isolation'); new combinations (*distanciamiento social* 'social distancing' or *distancia de seguridad* 'safety distance', *nueva normalidad* 'new normal', *barbijo quirúrgico* 'surgical mask', *aislamiento social* 'social isolation', *contacto estrecho* 'close contact'); or neologisms formed by derivation or composition (*anticuarentena* 'anti-quarantine', *desconfinamiento* 'unlockdown', *intrafamiliar* 'intra-family', *pospandemia* 'post-pandemic') (cf. Battaner 2021). There are also crude anglicisms (such as *home office*), often linked to new practices popularized in 2020.

In its 23rd edition, updated in December 2020,[10] the DLE features changes in the meanings of *confinado* 'confined' and *confinar* 'confine', while the entry for *confinado, da* 'confined' (masculine and feminine) was also amended. One new meaning of *confinamiento* 'confinement', related to 'lockdown', was added:[11]

---

8 The 24th edition of the DLE will differ from all previous editions and will involve a thorough overhaul of the work in a wide range of its structural elements. The new DLE "will be digital from its very conception. It is no longer a matter of converting into an electronic resource something that was conceived, and in part developed, as a work intended to be published on paper, but of creating a genuinely electronic dictionary, with everything such a fundamental fact implies" (https://www.asale.org/noticias/la-rae-presenta-la-primera-actualizacion-de-la-23a-edicion-de-su-dle).

9 The first update of the 23rd edition was done in December 2017; it was then updated again in December 2018 and in December 2019. This article focuses on the latest update, in December 2020. The academies are currently working on the 23rd edition and on the 24th edition simultaneously.

10 This update includes more than 2,000 changes.

11 The Fundación del Español Urgente (Foundation for Urgent Spanish), promoted by the EFE Agency and the Real Academia Española, even chose "confinamiento" as the word of the year 2020.

confinado, da. [Amended entry]. [. . .] adj. 1. Dicho de una persona: Obligada a vivir en un determinado lugar. U. t. c. s. ‖ m. y f. 2. Persona sometida a un confinamiento (‖ aislamiento impuesto a una población). ‖ 3. Der.En algunos países, persona que sufre la pena de confinamiento.

[(adjective) 1. Said of a person: Forced to live in a given place. Also used as a noun. 2. (Masculine and feminine) Person subjected to a confinement. (‖ isolation imposed on a population) ‖ 3. (Law) In some countries, a person who suffers a penalty of confinement.]

confinamiento. [Amended meaning] m. 1. Acción y efecto de confinar o confinarse. [(Masculine) 1. Action and effect of confining or confining oneself.]

confinamiento. [Meaning added].‖ m. 1 bis. Aislamiento temporal y generalmente impuesto de una población, una persona o un grupo por razones de salud o de seguridad. *El Gobierno decretó un confinamiento de un mes*.

[(Masculine) 1. bis Temporary and generally imposed isolation of a population, a person, or a group due to health or safety reasons. *The Government decreed a month-long confinement*.]

confinar. [. . .] ‖ 2. [Amended meaning]. tr. Encerrar o recluir algo o a alguien en un lugar determinado o dentro de unos límites. U. t. c. prnl. *Se confinó EN su casa*." [(Transitive verb) 2. Lock up something or someone in, or commit them to, a given place or within certain limits. Also used pronominally. *They confined themselves IN their home*.][12]

Cuarentenar 'quarantine' and its variants *cuarentenear* and *encuarentenar* were added as verbs in the DLE and one of the meanings of *cuarenteno* 'quarantined person' was amended:

cuarentenar. [Entry added]. tr. 1. Poner algo o a alguien en cuarentena (‖ aislamiento preventivo por razones sanitarias). *Cuarentenaron un hospital*. U. t. c. prnl. *Se cuarentenó durante la epidemia*. ‖ intr. 2. p. us. Pasar un período de cuarentena (‖ aislamiento preventivo por razones sanitarias). *Se permite el regreso a la ciudad de origen para cuarentenar*.

[1. (Transitive verb) To place something or someone in quarantine (‖ preventive isolation for health reasons). *They quarantined a hospital*. Also used pronominally. *They quarantined themselves during the epidemic*. ‖ 2. (Intransitive verb, scarcely used) To go through a period in quarantine (‖ preventive isolation for health reasons). *They are allowed to return to their home town to quarantine*.]

cuarentenear. [Entry added]. intr. 1. Pasar un período de cuarentena (‖ aislamiento preventivo por razones sanitarias). *Es más llevadero cuarentenear con alguien*.‖ tr. 2. p. us. Poner algo o a alguien en cuarentena (‖ aislamiento preventivo por razones sanitarias). *Tendremos que cuarentenear el ganado*. *Las autoridades cuarentenearon el crucero*.

[1. (Intransitive verb) To go through a period in quarantine (‖ preventive isolation for health reasons). *Quarantining is more bearable if you do it with someone*. ‖ 2. (Transitive verb, scarcely used) To place something or someone in quarantine (‖ preventive isolation for health reasons). *We will have to quarantine the livestock. The authorities quarantined the cruise*.]

---

12 https://www.rae.es/noticia/la-rae-presenta-las-novedades-del-diccionario-de-la-lengua-espanola-dle-en-su-actualizacion.

encuarentenar. [Entry added]. tr. Poner algo o a alguien en cuarentena (‖ aislamiento preventivo por razones sanitarias). *Si alguien se infecta, habrá que encuarentenar a toda la colonia*. U. t. c. prnl. *Me encuarentené por precaución*.

[(Transitive verb) To place something or someone in quarantine (‖ preventive isolation for health reasons). *If anyone is infected, the whole colony will have to be quarantined*. (Also used pronominally) *I quarantined myself just to be safe*.

cuarenteno, na. [. . .] ‖ 7. [Amended entry]. f. Aislamiento preventivo a que se somete durante un período de tiempo, por razones sanitarias, a personas, animales o cosas".

[‖ 7. (Feminine) Preventive isolation that persons, animals, or things are placed under for health reasons.][13]

Also, *desconfinamiento* 'unlockdown' (noun) and *desconfinar* 'unlockdown' (verb) were added as antonyms of *confinamiento* and *confinar*, respectively:

desconfinamiento. [Entry added]. m. Levantamiento de las medidas impuestas en un confinamiento.

[(Masculine). Lifting of the measures imposed in a confinement (or lockdown).]

desconfinar. [Entry added]. tr. Levantar las medidas de confinamiento impuestas a una población, o a parte de ella, en un territorio u otro lugar. U. t. c. intr. y c. prnl.

[(Transitive verb) To lift confinement (or lockdown) measures imposed on a population, or part of it, in a territory or other place. (Also used as intransitive and pronominally).][14]

New definitions were given to the entry *mascarilla* to adapt it to the meaning of 'mask':

mascarilla. [. . .] 2. [Amended meaning]. f. Máscara que cubre la boca y la nariz de su portador para protegerlo de la inhalación y evitar la exhalación de posibles agentes patógenos, tóxicos o nocivos. Mascarilla quirúrgica, sanitaria.

[(Feminine) 2. A mask that covers the mouth and nose of the wearer to protect them from inhaling and preventing them from exhaling possible pathogenic, toxic, or noxious agents. Surgical, sanitary mask.][15]

Coronavírico 'coronaviral' and *coronavirus* were added as terms from the field of medicine. It is interesting to note that an etymology for *coronavirus* was also added, indicating that it is derived from the English word *coronavirus*, but also recognizing Latin as the language from which it was derived originally.

coronavírico, ca. [Entry added]. adj. Med. Perteneciente o relativo al coronavirus.

[(Adjective, medicine) Of or relating to coronavirus.]

13 Ibid.
14 Ibid.
15 Ibid.

coronavirus. [Entry added]. m. Med. Virus que produce diversas enfermedades respiratorias en los seres humanos, desde el catarro a la neumonía o la COVID. [(Masculine, medicine) A virus that causes various respiratory diseases in humans, from the common cold to pneumonia or COVID.]

coronavirus.[Etymology added to the entry]. (Del ingl. *coronavirus*, de *corona* 'corona solar', por el aspecto del virus al microscopio, y este del lat. *corōna* 'corona', y *virus* 'virus', y este del lat. *virus* 'veneno', 'ponzoña').

[From the English *coronavirus*, from *corona* 'solar corona', because of the appearance of the virus under the microscope, and this from the Latin *corōna* 'corona' and *virus* 'virus', derived in turn from the Latin *virus* 'venom', 'poison'.][16]

The term *coronavirus* would appear to be a prototypical case of the model the Academy is adopting in an effort to address criticism it has received in the past. A brief definition is provided, which is simply phrased and straightforward. It includes the core propositions of the concept, so it can serve as a starting point for readers to work with texts of greater scope or depth, such as specialized texts.

This is an accurate definition, as it does not deny the fact that while it has gained visibility in the current pandemic, the term *coronavirus* has existed since 1968 (Ochoa Montes/Ferrer Marrero 2021), with seven varieties of the virus having been discovered thus far, including SARS-CoV-2. It also avoids the need to have an entry under the specific name of the virus that caused the pandemic.

At the same time, the definition of *coronavirus* mentions *COVID*, although the World Health Organization announced on February 11, 2020, that the official name of the disease would be *COVID-19*, a contraction of the term *coronavirus disease 2019*. However, in the DLE, *COVID-19* is recorded but it refers back to *COVID*, perhaps because it is assumed that that is the most frequent form used in the Spanish language.

A new entry was added for *COVID* as a medical term and its English etymology was also added. The DLE states that this word can be a feminine or masculine noun. In the process of its adaptation to different varieties of Spanish, it has developed two different accentuations: *covid* and *cóvid*, although this is not mentioned in the dictionary.

COVID. [Entry added]. m. o f. Med. Síndrome respiratorio agudo producido por un coronavirus. COVID-19. m. o f. Med. COVID.

[(Masculine or feminine, medicine) Acute respiratory syndrome produced by a coronavirus. COVID-19. (masculine or feminine, medicine). COVID.]

COVID.[Etymology added to the entry]. (Del ingl. *COVID*, y este acrón. de coronavirus *disease* 'enfermedad del coronavirus'.]

16 Ibid.

[From the English *COVID*, the acronym of *coronavirus disease*.][17]

The criteria that Bernal et al. (2020) had already observed in the 2019 update are also manifested in this update. Among these criteria, the most prominent in academic lexicographic practice is the consistency in updating a word with its derivatives. That lends rigor and internal consistency to the dictionary. In addition, a cautious attitude can be observed in the assessment of the frequency of a word and its stability. There is no urgency, at least in specialized language, to incorporate new words or meanings. As a result of the pandemic only two words with the label '*Med.*' were included: *coronavírico, ca* and *coronavirus*. As we will see in section 4, the DHLE takes more liberties.

3.2 General characteristics of the DHLE update

One of the most noteworthy developments for Spanish language lexicographic research is the availability, since 2013, of some entries of the *Diccionario histórico de la lengua española* (DHLE, 2013–). This is the result of an academic pan-Hispanic project, which received its final ratification with the establishment, in April 2021, of the Red Panhispánica de Academias, Universidades y Centros de Investigación para la Elaboración del Diccionario Histórico de la Lengua Española (Pan-Hispanic Network of Academies, Universities, and Research Centers for the Production of the Historical Dictionary of the Spanish Language, https://www.rae.es/dhle/).

According to its introduction, it is a "digital-native dictionary that seeks to fully describe (diatopically, diastratically, and chronologically) the history of the Spanish language lexicon," as well as to "analyze the history of the lexicon in a relational perspective, addressing the etymological, morphological, and semantic connections that are established between words" (DHLE 2013–). As a diachronically oriented database, the persons behind it aim to organize the entries by semantic fields or lexical families. It is important to note that the DHLE shares the same databases with the DLE, but it has its own seal and independent funding.

A detailed explanation of the DHLE's structure, its characteristics, and the type of relations between words and meanings (morpho-etymological, change mechanisms, semantic) can be found in a document that is available on its website. The document also clearly defines the terms of reference (lemma, sub-lemma, hyperlemma, meaning, sub-meaning, variant, syntactic scheme, documentation, multiword unit, and lexical family). In addition, four types of labels are identified for the meanings included in the DHLE: diatopical, pragmatic, sociolinguistic, and, most importantly for this study, diatechnical or specialization labels.

17 Ibid.

Available online with more than six thousand entries, the aim of the DHLE is to show, in an organized manner, the evolution of the Spanish lexicon over time and up to the present. In March 2021 alone, 715 new monographs were added, which led to the production of entries for words in different semantic fields and lexical families. A difference observed between the DLE and DHLE is that the former usually incorporates new words or meanings in more or less numerous batches, on a regular basis; whereas incorporations in the DHLE are done by lexical families, among other criteria.

The context of the pandemic has been particularly productive in terms of recording new expressions. A brief exploration of the DHLE revealed that some of these expressions are related to the virus, but most of them have to do with COVID-19, with its prevention and treatment, and with concepts that are only understood in the new circumstances. Some of these expressions are scientific in nature, but in other cases their inclusion in specialization languages could be debatable. This is the case of words such as *covidengue* 'covidengue' and *covidfobia* 'covidphobia', for example. Others seem to be humorous or somewhat picturesque productions. This is the case, for example, of a significant number of entries beginning with COVI-, such as *covicho*, 'covidbug', *covidcidio* 'covidcide', *covidiota* 'covidiot', and *covidemia* 'covidemia'.

With the aim of examining in greater detail a sample of expressions in vogue during the pandemic, which came to the attention of the DHLE and the DLE, as well as assessing their potential novelty, we chose to study a homogeneous universe: that of the words that emerged as derivatives or developments of CORONA-. This will be explained in the following sections.

4 The case of words beginning with CORONA-

4.1 Research design

As noted in the introduction, this section will present a concrete analysis whose universe is formed by the expressions beginning with CORONA-, provided they relate in some way to the Coronavirus. In considering only the terms that begin with CORONA-, the study leaves out many other words related to the pandemic, such as, for example, words beginning with COVI-, or words that are part of the set of words related to the prevention of the disease, such as *distancia social* 'social distance', *mascarilla* 'face mask' or *sanitización* 'sanitization'. In this sense, the universe studied constitutes a partial sample with respect to the total of words or meanings that can refer to the pandemic in the dictionaries studied.

The corpus includes: (i) the 23rd printed edition of the DLE, also known as the Tercentenary edition (DLE 2014); (ii) the 23.4 DLE version, which is currently online (DLE 23.4 2020); and (iii) the DHLE version available as of the date of this study, May 2021 (DHLE 2013–).

While the three works that make up the corpus are not completely homogeneous, neither in their structure nor in their objectives, for the purposes of this analysis the findings are comparable and allow for general and specific conclusions to be drawn regarding the recording status of neologisms, or neologism candidates, in the Spanish language in the pandemic scenario that unfolded as of 2019.

It should be noted that equivalents and variants, which are moreover only included in the DHLE, are recorded together or separately following the criterion used in the dictionary itself. In this sense, *coronasalmonela* and *coronasalmonella* each have their own entry because they are featured separately in the DHLE, but *coronadengue* and *corona-dengue* are considered as one entry because that is how they are recorded.

4.2 Research findings

The data presented below shows, as noted above, the recording status of words beginning with CORONA- and which are related to the coronavirus, in the three sources selected as corpus (DLE 2014; DLE 23.4 2020; DHLE 2013–).

The thirty-one words identified are *coronaplauso* 'coronapplause', *coronabebé* 'coronababy', *coronabicho* 'coronabug', *coronaboda* 'coronawedding', *coronabono* 'coronabonus', *coronabulo* 'coronahoax', *coronachikunguña* 'coronachikunguya', *coronachivato* 'coronainformer', *coronacompra* 'coronapurchase', *coronacrisis* 'coronacrisis', *coronadengue, corona-dengue* 'coronadengue', *coronadiccionario* 'coronadictionary', *coronadivorcio* 'coronadivorce', *coronafiesta* 'coronaparty', *coronafobia* 'coronaphobia', *coronahisteria* 'coronahysteria', *coronahistérico,a* 'coronahysteric', *coronalengua* 'coronalanguage', *coronalenguaje* 'coronalanguage', *coronamanía* 'coronamania', *coronacionalismo* 'coronationalism', *coronapositivo* 'coronapositive', *coronasalmonela* 'coronasalmonella', *coronasalmonella* 'coronasalmonella', *coronaviral* 'coronaviral', *coronavírico, a* 'coronaviral', *coronavirología* 'coronavirology', *coronavirólogo, a* 'coronavirologist', *coronavirosis* 'coronavirosis', *coronaviroso, a* 'coronainfected' and *coronavirus* 'coronavirus'. All thirty-one words that fit the search equation are recorded in DHLE, while only two of them are recorded in DLE 23.4 (*coronavírico, ca* and *coronavirus*), and none in DLE 23. Thus, the only two words beginning with CORONA- that are featured both in DLE 23.4 and DHLE are *coronavirus* and *coronavírico, ca*.[18] None of the humorous constructions, such as *coronaplauso*, *coronabebé*, or *coronabicho* are recorded in DLE, perhaps in the understanding that their stability is not ensured.

Four of the thirty-one words included in DHLE were documented before the pandemic (*coronavirus*, since 1980; *coronavirosis*, since 1992, *coronaviral*, since 1997; and *coronavirología*, since 2012). This thus verifies that, while these documented instances

18 The entries for these two words in both dictionaries are attached as figures 1–4 in annex I.

predate 2014 (year in which the DLE 23rd edition came out), at that time there were not sufficient arguments to include any of those words in that dictionary, or in its 23.4 online edition. Documented instances may exist prior to 2019 because *coronavirus* is a generic term coined in 1968. In fact, a research study conducted in the United Kingdom toward the mid-1960s revealed that the first of what are now known as *coronaviruses* corresponded to a virus found in chickens suffering from bronchitis, around the year 1930 (Tyrrell/Bynoe 1965). SARS-CoV-2 is just one of the viruses in the Coronavirus family. This clarification is necessary to determine which words are neologisms or potential neologisms, and which have a long-standing existence, although without the exposure they have now.

Of the thirty-one words recorded in DHLE, only two have a diatechnical label, and in both cases, it is medicine ('*Med.*'). *Coronavirus* receives a diatechnical label both in the DLE 23.4 and DHLE, while *coronaviral* has one but only in DHLE, given that the word is not recorded in DLE or in DLE 23.4. The DHLE also assigns it a double label: medicine and veterinary. The adjective *coronavírico, ca*, for its part, has the '*Med.*' label in the 23.4 edition of DLE but not in DHLE.

There are at least four other expressions, featured only in DHLE, that could be considered specialized words that merit a diatechnical label, although they do not have one, at least to date: *coronavirología*, *coronavirólogo*, *coronavirosis*, and *coronaviroso*. These four words have not gone beyond their specialized field and do not appear to be used outside it, so that they meet the generally accepted criteria for receiving a diatechnical label.

Coronabono, for its part, is a word clearly connected with the field of economics, an area of knowledge that has its own diatechnical label in DLE, and which could constitute another case to consider.

The twenty-seven words in DHLE that have no documented instances prior to the pandemic could be considered candidates for full neologisms, taking into account also that they were not recorded in the last 23.4 version of DLE.

None of the dictionaries incorporates any foreign words connected with the pandemic.

In Table 1, the thirty-one words identified are distributed geographically based on the data recorded only in DHLE, given that that dictionary is the only one of the three sources that features all the words, and also locates each documented instance geographically. This table also has a chronological reference, as DHLE indicates the year of the documented use.

The geographical indications are divided into four large regions: Europe, North America, Central America and the Caribbean, and South America. This division, however, precludes the drawing of reliable conclusions or trends regarding the scope of use. Moreover, as previously noted (Bernal et al. 2020), applying exclusively a frequency criterion is not enough in the usual practice for updating language dictionaries, so that the geographical data gathered here has illustrative rather than descriptive value.

Table 1: Geographical and chronological distribution of the documented uses of the thirty-one corona- words in DHLE.

Word	**Europe**	**North America**	**Central America and the Caribbean**	**South America**	**Total**	**2020**	**2021 (as of 30/05)**
Coronaplauso	4				4	2	2
Coronabebé				6	6	6	
Coronabicho	4	3			7	6	1
Coronaboda	3			2	5	4	1
Coronabono	6		2		8	8	
Coronabulo	11				11	9	2
Coronachikunguña				1	1		1
Coronachivato	2				2	1	1
Coronacompra	1	2	1	1	5	4	1
Coronacrisis	4			6	10	8	2
Coronadengue		2		5	7	4	3
Coronadiccionario				1	1	1	
Coronadivorcio	2	2		1	5	5	
Coronafiesta	2	2		2	6	4	2
Coronafobia	5		1	2	8	6	2
Coronahisteria	5	1			6	5	1
Coronahistérico,a	2				2	2	
Coronalengua	1			1	2	2	
Coronalenguaje	1	1			2	2	
Coronamanía	1			1	2	2	
Coronacionalismo	4				4	4	
Coronapositivo,a				4	4	4	
Coronasalmonela				1	1		1
Coronasalmonella				1	1		1
Coronavírico,a	7	1			8	6	2
Coronavirólogo,a	4	1		1	6	4	2
Coronaviroso,a	1		4		5	4	1
TOTALS	70	15	8	36	129	103	26

Only one word (*coronacompra*) is recorded in every region, which does not mean that it is used in every country of those regions.

While the criteria for the inclusion/exclusion of neologisms in DHLE are not made explicit, an inductive analysis of the recorded instances of the thirty-one words studied allows us to conclude that only one documented instance is needed in order to be included in DHLE, as in Table 1 there are four cases with a single documented instance (*coronachikunguña*, *coronadiccionario*, *coronasalmonela*, and *coronasalmonella*).

The maximum number of documented instances is eleven, and there is only one word with that many instances. This may reflect the very recent emergence of all of these words, in line with the unfolding of the pandemic. Moreover, of the 129 documented instances in total, 103 are from the year 2020 (80 percent) and 26 (20 percent) from the year 2021, although the analysis for 2021 only goes up to May 30.

While we understand the logic that is intended to be established for the progressive construction of DHLE, as far as neologism candidates are concerned, with this repertoire, in addition to acting as a historical dictionary, it also becomes, in practice, a refined emergency dictionary[19] that gathers new words – even some whose validation could be questionable – that in the future could either be potential lemmas of DLE or discarded outright for not meeting the usual criteria for incorporation into the main dictionary of the Spanish language.

5 Final considerations

A pandemic of the magnitude of the one that broke out in 2019, and with the speed at which it has spread around the world, is unusual. With the virus, the disease, and the social and economic changes brought on by the pandemic, words and terms also spread very rapidly, quickly capturing the attention of linguists and lexicographers. The academic dictionaries of the Spanish language are facing this situation in an unprecedented technological and methodological scenario for Spanish academic lexicography, which allows for an aggiornamento never seen before with respect to the updating practices both for DLE and DHLE. Two absolutely exceptional situations thus converged: the pandemic and a modern updating system.

At the same time, as the possibility of updating the DLE coincides with a momentum in the DHLE, the academic outlook is even more unique. DHLE operates

19 This term stands for the Spanish phrase “diccionario de emergencia” or “diccionario de urgencia” (emergency or urgency dictionary), as it is used on websites such as http://www.intranet.senasa.gov.ar/intranet/imagenes/archivos/prensa/caja_herramientas/Diccionario_de_Urgencia.pdf or https://www.meneame.net/m/actualidad/mami-guillao-masacote-diccionario-urgencia-descifrar-canciones or https://www.consumer.es/economia-domestica/finanzas/diccionario-de-urgencia-para-entender-que-ocurre-con-los-bancos-espanoles.html (last access: 10 June 2022).

methodologically as an emergency dictionary; it does not abandon its role as historical dictionary, but it acts also as an emergency dictionary, while DLE functions as an exclusion corpus. DHLE plays the role of witness corpus, in which all lexical creations are recorded. By recording neologisms it is writing the history of these words, bearing witness to terms that may or may not remain in the language, but whose appearance and disappearance can be dated and documented. DLE will incorporate only some of these words added by DHLE. It will incorporate only those that have stability, that can still be observed after a reasonable period, that have resisted the passage of time. All the words that remain in DHLE and are not incorporated into DLE are in quarantine, or, more precisely, in limbo, because it is a quarantine that is not necessarily going to end. They may remain there for an indefinite period of time, unless usage determines otherwise.

In this particular aspect, DHLE can be compared to the *Diccionario manual e ilustrado de la lengua académica* (1989) (Manual and Illustrated Academic Language Dictionary),[20] in that it distinguishes, as Alvar noted (1992), the neologisms-general/usual language dictionary relationship from the neologisms-manual dictionary relationship. Alvar understands that the manual dictionary gathers new words "aware that they could be a vocabulary that will have a fleeting existence in the general language." And he adds, "this is a necessary process: these words may disappear without leaving any other trace than the ephemeral presence of a limited use, but they may become widespread in their use and this non-normative repertoire will have been the anteroom for accessing the *Diccionario usual* [DLE]" (Alvar 1992). The idea of an "anteroom" can also be applied to DHLE.

Moreover, it should be noted that in DLE there is a reluctance to incorporate words, because there is a clear awareness of the difficulty involved in removing a word from the dictionary once it is included. In contrast, DHLE does not face a horizon in which it will be necessary to discuss whether or not a word is removed from the dictionary: it only needs to document the year of its last recorded use.

DLE aims, as it should, to be increasingly more user-friendly for native speakers, who, incidentally, are its intended audience; DHLE, by dating the first documented instance of the word in question, becomes a very user-friendly dictionary for researchers. Researchers often find it difficult to see in DLE the principles and criteria that govern the incorporations into the dictionary. These principles and criteria are not made explicit with the 2020 incorporations either, but it is clear that it has adapted to the emergence of the pandemic and that it took advantage of the technological resources available, which are, moreover, part of its new institutional policies.

The analysis of the nature of the new pandemic-specific expressions reveals that only two words in DLE (*coronavirus* and *coronavírico, ca*) are, strictly speaking,

---

20 Cf. also *Diccionario esencial de la lengua española* (2006).

specialized terms: they receive a modern lexicographic treatment, which does not pose past difficulties when it comes to defining specialized language words.

These two words are also recorded in DHLE, but only *coronavirus* has the label '*Med.*' As noted, other words beginning with CORONA- could be assigned that label too.

The new additions open a debate on the treatment of neologisms in lexicography, in a particularly unique scenario. It could be said that the changes made in DLE – as a result of the coronavirus pandemic – in a way rekindle old discussions regarding the criteria and methods used by DLE to select, incorporate, and define expressions belonging to specialized areas.

The pandemic represents an opportunity for lexicography and terminology researchers to discuss and propose consistent solutions for the incorporation of scientific and specialized words into DLE and other Spanish dictionaries. In this regard, it offers a chance to leave behind vague criteria for incorporating or excluding scientific terms, scientific definitions not easily understood by a regular audience, conceptual inaccuracies, and somewhat erratic assignments of thematic labels, among other criticisms that DLE has received.

Annex I

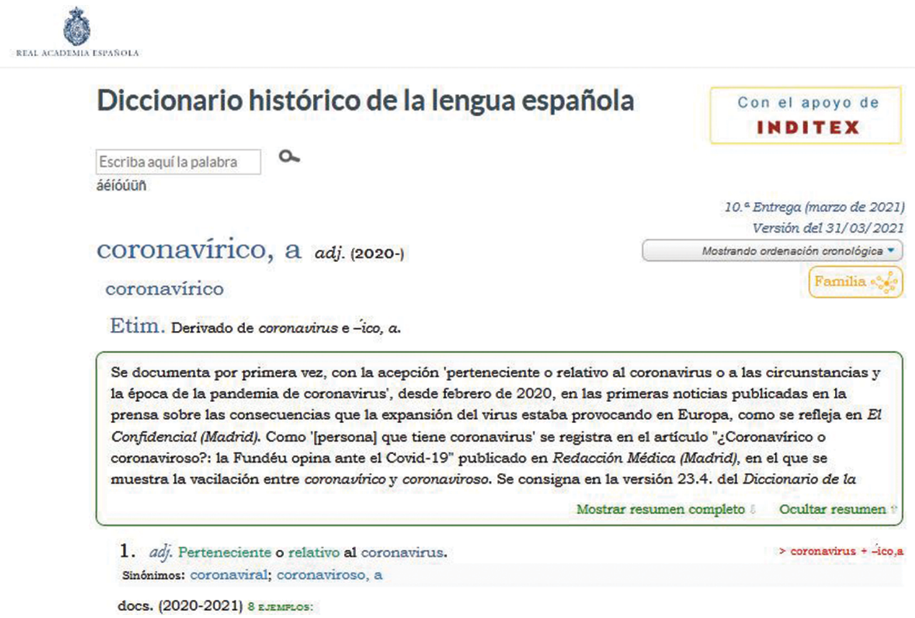


Figure 1: https://www.rae.es/dhle/coronavírico.



Figure 2: https://www.rae.es/dhle/coronavirus.



Figure 3: https://dle.rae.es/coronavirus.



Figure 4: https://dle.rae.es/coronavírico.

Bibliography

Adelstein, Andreina/Freixa, Judit (2013): Criterios para la actualización lexicográfica a partir de datos de observatorios de neología. Conference paper presented at *Congreso Internacional El Diccionario: neología, lenguaje de especialidad, computación*, October 28–30, 2013. Mexico City, Mexico. [https://repositori.upf.edu/handle/10230/34891, last access: July 1, 2021].

Alvar, Manuel (1992): Los diccionarios académicos y el problema de los neologismos. In: Reigosa, Carlos G. (coord.): *El neologismo necesario*.Madrid: Fundación EFE, 51–70.[<http://www.cervantesvirtual.com/obra-visor/los-diccionarios-acadmicos-y-el-problema-de-los-neologismos-0/html/01099ca2-82b2-11df-acc7-002185ce6064_6.html#I_0_>; last access: July 1, 2021].

Barité, Mario/Blanco, Amanda (2014): Teoría y pragmática de las marcas temáticas: estudio comparativo en diccionarios de lenguas española y portuguesa. In: *I Simposio Interdisciplinario de Organización del Conocimiento y Lexicografía: caminos que se cruzan, Montevideo, Universidad de la República, 9 y 10 de octubre de 2014*. Montevideo: FIC.

Battaner, María Paz (1996): Terminología y diccionarios. In: *Jornada Panllatina de Terminologia*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada, 93–117.

Battaner, Paz (2021): Cambios léxicos que ha generado la pandemia de la COVID en el mundo hispanohablante. Conference presented at the Real Academia Española, Instituto de España. 1 de junio 2021.

Bernal, Elisenda/Freixa, Judit/Torner, Sergi (2020): Criterios para la diccionarización de neologismos: De la teoría a la práctica. In: *Revista Signos: Estudios de Lingüística* 53(104): 592–618. [https://www.scielo.cl/pdf/signos/v53n104/0718-0934-signos-53-104-592.pdf, last access: July 1, 2021].

DHLE = *Diccionario histórico de la lengua española*. Ed. by Real Academia Española. Madrid: RAE, 2013– . . .. [https://www.rae.es/dhle/, last access: July 1, 2021].

Diccionario histórico de la lengua española: Estructura y características de los artículos del Diccionario. Madrid: RAE, n.d. [https://www.rae.es/dhle/docs/DH_estructura_2020.pdf, last access: July 1, 2021].

Diccionario manual e ilustrado de la lengua española. Ed. by Real Academia Española. 4th ed. Madrid: Espasa-Calpe, 1989.
DLE = *Diccionario esencial de la lengua española*. Ed. by Real Academia Española y Asociación de Academias de la Lengua Española. Madrid: Espasa, 2006.
DLE = *Diccionario de la lengua española*. Ed. by Real Academia Española and Asociación de Academias de la Lengua Española. 23rd ed. (Edición del Tricentenario). Buenos Aires: Espasa, RAE, and ASALE, 2014.
DRAE 2001 = *Dicccionario de la lengua española*. Ed. by Real Academia Española. 22nd. edition. Madrid: Espasa.
DLE 23.4 = *Diccionario de la lengua española*. Ed. by Real Academia Española and Asociación de Academias de la Lengua Española,23.4 online ed. Madrid: RAE and ASALE, 2020. [https://dle.rae.es/, last access: July 1, 2021].
Martínez de Sousa, José (2009): *Manual básico de lexicografía*. Gijón: Trea.
Moreno Moreno, María Águeda (2019): Términos entre la ciencia y la tradición: diccionario y sus sinónimos. In: *Lengua y Habla* 23, 76–90.
Ochoa Montes, Luis A./Ferrer Marrero, Daisy (2021): Daño cardiovascular en la COVID-19: Una extensión de la enfermedad pulmonar. In: *CorSalud* 13(1),68–85. [http://www.revcorsalud.sld.cu/index.php/cors/article/view/715/1298, last access: July 1, 2021].
Simen, René Tondji (1996): *Problématique des marques d'usage dans des dictionnaires monolingues et bilingues du français et de l'anglais: Thèse de doctorat*. Montréal: Université de Montréal. [http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp05/nq21521.pdf, last access:July 1, 2021].
Tyrrell, D. A. J./Bynoe, M. L. (1965): Cultivation of a novel type of common cold virus in organ culture. In: *British Medical Journal*/ 1, 1467–1470.

Judit Papp

How the COVID-19 pandemic is changing the Hungarian language: Building a domain-specific Hungarian/Italian/English dictionary of the COVID-19 pandemic

1 Introduction

This paper presents the main issues connected with the creation of a trilingual Hungarian-Italian-English dictionary of the COVID-19 pandemic using Lexonomy.[1] My aim is not only to create a coronacorpus (in Hungarian, I propose my own corona-neologism or 'coroneologism':[2] *koronakorpusz*) and a dictionary of equivalents, but also to understand how the different waves and phases of the COVID-19 pandemic are changing the Hungarian language, detect the *Corona-*, *COVID-*, *pandemic-*, *virus-*, *mask-*, *quarantine-*, and *vaccine*-related neologisms, and offer an overview of the most frequent or linguistically interesting Hungarian neologisms and multiword units related to COVID-19.

For the creation of the Hungarian/Italian/English dictionary of the COVID-19 pandemic (hereinafter referred to as the *Trilingual (HU, IT, EN) COVID-19 Dictionary*, TCD), I used a specialized *coronacorpus* extracted from the Web using Sketch Engine.[3] To detect the related terms, I also analyze the Hungarian web corpora of news articles (online press) obtained from crawling a list of RSS feeds (Timestamped JSI web corpus).[4] It is already highly evident that the vocabulary used in these articles (in the printed versions as well as in online press and media) is rather different with respect to the past. In fact, it is possible to note a frequency increase (for a short period, such as from March to the end of May 2020, or for a longer period, such as from March to the end of 2020) for certain word forms that are to some extent related to the all-encompassing COVID-19 pandemic. It is also possible to discover word forms that, before the outbreak of the pandemic, have never been seen in everyday

1 https://www.lexonomy.eu/p8mwspck/ (last access: 10 June 2022).
2 The COVID-19 inspired neologisms or 'coronacoinages' are sometimes referred to also as 'coroneologism', e.g., in papers written by Roig-Marín (2020). Previously, the term 'coroneologism' appeared in newspaper articles (e.g. Coroneologisms are going viral. In: *Economic Times*. April 9, 2020).
3 https://www.sketchengine.eu/ (last access: 10 June 2022).
4 https://www.sketchengine.eu/jozef-stefan-institute-newsfeed-corpus/ (last access: 10 June 2022).

Judit Papp, Department of Literary, Linguistic, and Comparative Studies, Palazzo Santa Maria Porta Coeli, University of Naples L'Orientale, Italy. Via Duomo 219 – 80138, Naples, Italy.
e-mail: jpapp@unior.it

https://doi.org/10.1515/9783110798081-008

Hungarian language. Many terms that usually belong to the medical and scientific fields (*epidemiology*, *virology*, *serology*, etc.) are being used in everyday language (in the press but also in informal contexts).

For the domain-specific terminology extraction, I used the Oneclick Dictionary function of Sketch Engine and created the first drafts of TCD.

From the dictionary drafts, I extracted the headwords related to the pandemic and included them in the TCD. I customized the structure and formatting of the dictionary in Lexonomy as well as configured the connection with my Sketch Engine account to have the possibility to extract and pull example sentences from Sketch Engine.

Finally, I completed the entries with the Italian and English equivalents and the corresponding examples taken from the Web as well as from the corresponding Timestamped JSI web corpora.

2 Field of study

Studies and research dedicated to the methodical lexicographic treatment of Hungarian terms related to the COVID-19 pandemic are still rather uncommon. The existing glossaries or dictionaries are usually monolingual (Hungarian) or bilingual (Hungarian-English). Except for my own *Trilingual (HU, IT, EN) COVID-19 Dictionary* (TCD), there is no other existing Hungarian-Italian dictionary on COVID-19.

Following the first wave of the pandemic, in 2020, a dictionary on the lexicon of COVID-19 was published in Hungary by Ágnes Veszelszki, the *Karanténszótár*, which collects 400 neologisms (words and expressions) that have appeared in the Hungarian language between January and July 3, 2020. Each lemma is accompanied by an explanation and examples taken from real texts. Besides the most commonly used words and expressions, Veszelszki has also included rather rare forms as well as *hapax legomena* in her dictionary. The dictionary, accompanied by a short essay, is an authentic and important testimony of the period under review, as it offers users a detailed view of the Hungarian linguistic aspects of the COVID-19 pandemic. In addition, it constitutes a valuable source for further linguistic reflections on the formation of neologisms in the Hungarian context. The essay is also interesting for the lexicon used by Veszelszki, as it is profoundly influenced by the pandemic and has its own related neologisms, namely *karanténszótár* 'dictionary on quarantine', *karanténszókincs* 'lexicon on quarantine', *kórlenyomat* 'imprint of the disease', and *karanténkor* 'period of quarantine'. These neologisms appear either in the title, introduction, or essay of the publication, but the author does not lemmatize nor define them among the items collected.

Another noteworthy publication on the challenges posed by COVID-19 and the various responses to the pandemic is *Globális kihívás – lokális válaszok* (Global challenge – local responses) edited by László Kovács (2020), which includes a section

dedicated to the articles that reflect on the new phenomena in the Hungarian language (Balázs; Domonkosi-Ludányi; Kegyes- Lanzmaier-Ugri and Lénárt).

With the creation of the TCD, my aim has been to fill this lexicographic gap primarily concerning the Hungarian-Italian language pair and to organize this content in a free online tool (a rich database) that is easy to search and useful for linguists and translators. The dictionary created with Lexonomy makes it possible to store, maintain, and update data in an organized manner. The third language is English, as the comparison with it is inevitable. On the one hand, that is due to the enormous quantity of news produced and conveyed, with extraordinary speed, by international agencies, a phenomenon exerts a considerable influence on other languages and, on the other hand, because English is the language of the international scientific community including, therefore, international medical research. The papers, findings, and results of scientists' experiments relating to COVID-19 are published in English, which means that English plays an important role in the creation of neologisms. In both Hungarian and Italian, we record a certain number of loans, calques, and adaptations, but we also have to deal with the needs of ordinary people and the creative abilities of individual languages.

3 Methodology

3.1 Corpus Selection

The COVID-19 Open Research Dataset (CORD-19) that is available on Sketch Engine consists of a collection of texts in English. As of November 2021, I still cannot find any specific Hungarian or Italian COVID-19 related corpus or a Hungarian-Italian COVID-19 related dictionary.

Therefore, for creating TCD, I decided to build my own COVID-19 related Hungarian corpus (in Hungarian *koronakorpusz*) using Sketch Engine and starting from the Web, as it represents an enormous resource ('web as corpus', cf. Kilgarriff 2001, Kilgarriff-Greffenstette 2003). The Hungarian coroneologisms and words related to the COVID-19 pandemic are detected in this specific corpus that is built using all three options of Sketch Engine that make it possible to make the corpus larger:

- content downloaded by providing some typical words defining the topic (seed words),
- content downloaded by providing a list of URLs for download,
- content downloaded by downloading a complete website.

Content downloaded by providing the typical words that define the topic (seed words)
("Find texts on the Web" option, input type: "Web search")

As a result of this option, Sketch Engine extracted a series of web pages and documents. In "Web search", I input groups of words and phrases (maximum 20) to enable defining the topic of the new corpus. With the pandemic's progression and the succession of the different phases and waves, among others, I used seed words such as: *Astrazeneca*, *átoltottság* 'vaccination coverage rate', *COVID*, *COVID-19*, *COVID-igazolvány* 'COVID certificate', *delta*, *deltavírus* 'delta variant', *digitális* 'digital', *fertőtlenítés* 'disinfection', *fertőzés* 'contagion, infection', *fertőzött* 'infected', *görbe* 'curve', *harmadik* 'third', *hullám* 'wave', *immunitás* 'immunity', *járvány* 'epidemic', *karantén* 'quarantine', *koronavírus* 'coronavirus', *Moderna*, *mutáció* 'mutation', *oltás* 'vaccination', *oltásellenes* 'anti-vax', *oltáspárti* 'pro-vax', *oltópont* 'vaccination point', *Pfizer*, *Sputnik*, *szájmaszk* 'mask', *távolságtartás* 'social distancing', *tömeges* 'massive', *vakcina* 'vaccine', *vakcinabeszerzés* 'vaccine procurement', *védettség* 'immunity', *védőoltás* 'vaccine', *vírusvariáns* 'virus variant', etc.

An advantage of Sketch Engine is making it possible to run the corpus building tool many times to make it increasingly larger. It is also possible to repeat the search with the same seed words multiple times but also with different seeds, as well as to have multiword expressions using the quotes or proper names of different kinds. These words (seeds) are randomly selected and groups of three are sent to the Bing search engine. The Web pages that Bing returns are downloaded and processed into a corpus.

Content downloaded by providing a list of URLs that should be downloaded
I have also collected Hungarian language data from relevant URLs (e.g. blogs, forums, general websites on COVID-19, etc.). The main criterion for inclusion in the corpus is texts dealing with topics related to the pandemic.

Content downloaded by downloading a complete website
In particular, I have downloaded a few websites (July 12, 2021) containing useful information on the topic.[5]

---

5 The downloaded websites are: (i) https://koronavirus.gov.hu/, the official governmental portal in Hungarian on COVID-19 created together with the Operational Force responsible for the Prevention of the COVID-19 pandemic (*Koronavírus-járvány Elleni Védekezésért Felelős Operatív Törzs*) on January 31, 2020. The Ministry of Interior of Hungary is responsible for the operation of the portal and the Prime Minister's Office is responsible for editing the content; (ii) https://www.covid1001.hu/: in the middle of March 2020, a group of medical translators (specialists, biologists, pharmacists, epidemiologists, language specialists) decided to combat misinformation by translating and publishing reputable articles; (iii) https://semmelweis.hu/koronavirus/, Semmelweis University's website on the novel coronavirus, which is constantly updated with the latest news, information, communications, instructions, and actions concerning university citizens; (iv) https://www.elte.hu/content/koronavirussal-kapcsolatos-tajekoztatok-cikkek.c2c.316: Eötvös Lóránd University's website that contains information on COVID-19 (updated: June 30, 2021);

With the help of its corpus building tools, Sketch Engine processed many Web pages and documents and built the Hungarian ‘coronacorpus’ (about 4 million words).

For the domain-specific terminology extraction, I used the Oneclick Dictionary function of Sketch Engine and created the first drafts of TCD. The Oneclick Dictionary is useful in automating the exchange of lexicographic data between the selected Sketch Engine corpus and a Lexonomy dictionary (Měchura 2017), even if post-editing is required. Besides my own specialized corpus, I analyze the following Hungarian web corpora of a news articles obtained from crawling a list of RSS feeds: *Timestamped JSI web corpus 2014–2020 Hungarian* and *Timestamped JSI web corpus 2021–01 Hungarian*. “The JSI Newsfeed corpus is a new family of Web corpora created from the JSI newsfeed of Jozef Stefan Institute, Slovenia [. . .]. JSI newsfeed is a clean, continuous, real-time aggregated stream of semantically enriched news articles from RSS-enabled sites across the world.” (Bušta et al. 2017). The corpora are tagged by TreeTagger v2.

Concerning *Timestamped JSI web corpus 2014–2020 Hungarian*, I have also created a sub-corpus that contains only articles from 2020, including 309,663,951 tokens and 256,156,393 words. The *Timestamped JSI web corpus 2021–01 Hungarian* contains 113,132 documents, including 34,378,246 tokens, 28,376,390 words, 1,624,519 sentences, and 699,713 paragraphs. Although these corpora are obviously not exhaustive, given these figures and the wide coverage of Hungarian language sources, I conclude that the size of the corpora can be suitable for analyzing the phenomena and trends in the Hungarian online press.

While my *coronacorpus* is useful in detecting the Hungarian *coroneologisms* (accepted by the speech community) and occasionalisms (or ‘nonce words’ coined for a particular occasion, e.g. *aranymaszk* ‘gold mask’) used not only in newspaper articles and standard Hungarian texts (everyday, neutral, unmarked) but also on other websites (government, homepages, school/university, etc.), blogs, and social networks (Facebook, Instagram, etc.). In this three-way, colloquial language (slang, informal, familiar) and formal language (scientific, specialized, academic, literary) will also be represented in TCD.

The Hungarian Timestamped JSI web corpora are an outstanding tool to detect the behavior of the words or single word forms. ‘Trends’, in fact, is a feature of Sketch Engine “for detecting words that undergo changes in the frequency of use in time (diachronic analysis). Trends identify words whose use increases or decreases in time.”[6]

Alongside this feature, ‘Concordance’ is useful, mostly the ‘Distribution of hits in the corpus’ function provides highly informative results. The ‘Word Sketch’ option, a

(v) https://europa2000.hu/covid-19/. The COVID-19 section of the website operated by the Europa 2000 Secondary School (Budapest), which is a secondary grammar school and vocational institution maintained by a foundation; and (vi) https://www.pfizer.hu/, the Hungarian version of the institutional site of Pfizer, one of the world’s premier innovative biopharmaceutical companies.

6 3https://www.sketchengine.eu/guide/trends/#toggle-id-6.

one-page summary of the word's grammatical and collocational behavior, is another helpful feature (active in my own corpus, not available in the Timestamped corpora).[7]

A good example for illustrating the 'Trend' feature is the Hungarian coroneologism *nyunyóka*. The explosive growth of its frequency is strictly related to the pandemic. A *nyunyóka* can be anything that is safe for a baby or toddler to have at sleep time. It is a sort of comfort or transitional item – a blanket or stuffed animal or other comfort object of affection that a baby or toddler brings to bed, and that provides comfort and soothing. Previously, the term *nyunyóka* was uncommon and was used only in baby talk, and then, due to the massive media impact of Chief Medical Officer Cecília Müller's discourse during a press conference of the Operational Force, concerning personal hygiene habits to teach kids and the necessity to wash comfort objects frequently, this neologism entered the common language and became widely known and used. The number of hits found in the corpus is 166, for a lemma present only since May 13, 2020.[8] Müller shared these tips instead of the daily COVID numbers, mortality and recovery rates, current active cases, recoveries, etc. that people were actually expecting. The results of the search query using Google now list 46,500 pages (as of December 16, 2021).

Besides the Timestamped JSI web corpora, the Web is a valuable corpus to find coroneologisms and forms belonging to Hungarian slang or to the colloquial register. The latter forms are usually not represented in current corpora typically based on news articles, which is why the creation of the Hungarian *coronacorpus* is important for this research.

From these drafts, I extracted the headwords related to the pandemic and included them in TCD. I customized the structure and formatting of the dictionary in Lexonomy and configured the connection with my Sketch Engine account so that there is an option to extract and pull example sentences from it. This option allows you to detect, select, and pull not only definitions and descriptions of the Hungarian coroneologisms (new words, new meanings of existing words, and new multiword units) into Lexonomy, but also collocates and collocations, etc. While building the dictionary, particular attention is paid to neologisms related to aspects regarding the outbreak of the pandemic, lockdowns, curfews, quarantines, social distancing, good hygiene practices, epidemiological curves, smart working, distance learning, first

7 However, in 'Show visualization', it would be great if the image could be editable by the user.

8 From Müller's discourse (https://index.indavideo.hu/video/Csenjuk_el_a_gyermek_nyunyokajat): "Tudjuk jól, hogy a piciknél van valamiféle ragaszkodás: itt nemcsak a cumikra gondolok, hanem kis pelenkára, vagy nyunyókára, amit ő otthonról hoz és nagyon szereti. Próbáljuk meg ezeket otthon gyakran tisztítani, elcsenni ameddig alszik a gyermek és ezeket kimosni és vasalással még egy hőkezelésnek alávetni, ami szintén fertőtlenítő hatású." (We know very well that little babies have some kind of attachment. Here, I am thinking not only of the pacifiers, but also of the little diaper or any comfort object he or she brings from home (to the nursery) and loves it very much. Let's try to clean them frequently at home, sneak them away from the child while they are asleep and wash them and subject them to heat treatment with ironing, which also has a disinfectant effect.)

wave, second wave, burden on healthcare systems, vaccines, and vaccine efficacy, third wave, fourth wave, variants, green pass, and the EU digital COVID-19 certificate. These aspects are where the largest part of new words came into existence. Common Hungarian terms that are important for understanding the COVID-19 pandemic are also included in the dictionary.

As far as the Italian and English equivalents are concerned, I proceed with interrogating the available Timestamped JSI web corpora for Italian and English and the above-mentioned COVID-19 Open Research Dataset (CORD-19): *Timestamped JSI web corpus 2014–2020 Italian*, *Timestamped JSI web corpus 2021–01 Italian*, *Timestamped JSI web corpus 2014–2020 English*, *Timestamped JSI web corpus 2021–01 English* and the *COVID-19 Open Research Dataset* (CORD-19).

3.2 Terminological extraction

To be able to extract more and more COVID-19-related terms with my COVID-19-related Hungarian corpus and the Timestamped JSI web corpora, I used the 'Keywords' function (terminology extraction) that is available on Sketch Engine, downloaded and analyzed the 'Wordlist' (frequency list), and used the 'Concordance' function. In particular, in 'Wordlist' (BASIC tab), I searched for certain strings, such as *COVID*, *korona*, *karan* (from *karantén* 'quarantine'), *járvány* ('pandemic'), *vírus*, *fert* (from *fertőz* 'infect'), *beteg* ('ill'), *véd* ('protect'), *olt* ('to vaccinate'), *vakcina* ('vaccine'), and *immun* to get the productivity of the corresponding lemmas.

3.3 Draft dictionary and formatting

Again, with my own COVID-19 related Hungarian corpus, I used the One-Click dictionary (automatic dictionary drafting) function of Sketch Engine to create my draft dictionary for Hungarian. The result was useful. From the draft, I extracted many headwords related to the pandemic. However, after a while, I learned how to use Lexonomy as well as how to configure and customize the dictionary structure and formatting, then I preferred to create a new, empty dictionary using the 'Create a dictionary' option and insert data manually, one by one. This method is time-consuming, but the content is more professional. I have also configured the connection with my Sketch Engine account to connect TCD with one of the corpora and implement the information available for the single terms or expressions.

3.4 The structure of TCD

Headwords consist not only of single words, but they also include particularly frequent or relevant multiword expressions (MWEs). Less frequent MWEs are presented as collocations of the headword or among the examples.

TCD is linked to the original corpus in Sketch Engine, and it is possible to detect, select, extract, and automatically pull definitions, examples of usage, collocations, and thesaurus items of the Hungarian coroneologisms from my corpus into Lexonomy.

Common Hungarian terms that are important for understanding the COVID-19 pandemic are also included in the dictionary, such as *járvány* 'epidemic', *vírus* 'virus', *fertőzött* 'infected', etc.

The Italian and English equivalents are added manually along with useful examples taken from texts on the Web.

4 First results

The use of several pre-existing occasional words and expressions has increased significantly during the COVID-19 period, while neologisms linked to the pandemic were coined with surprising speed (e.g. *covidiot*, *coronababy*, *zoom-kocsma* 'virtual pub in Zoom', *fotelvirológus* 'armchair virologist', etc.).

The lexical innovation resulting from the explosion of the pandemic is incomparable, as terms inspired and/or linked to COVID-19 entered the large-scale public consciousness. Faced with the new reality, the neologisms represent a functional tool to discuss all of the different phenomena related to the pandemic: the impact that the pandemic and the crisis have on our lives, society, and economy, the experiences following restrictive lockdown measures, and the many themes related to distance learning or vaccines. They are also useful for expressing our feelings or making light of our experiences.

While, on the one hand, words and expressions that have dominated the pandemic-related discourse since the outbreak of the pandemic have an informative function, on the other hand, in a certain sense they also allow us to gain mutual understanding, to protect each other, to share warnings, to comment on events, to express and share with others anxieties, fears, worries, anger or exasperation. With their help, we can also make jokes, laugh, or make fun of this shared lexicon or even rid ourselves of fears. For this purpose, my COVID-19 related Hungarian corpus can be useful, as it contains language data also from social media sites, forums, and blogs. It such texts we register some of the most common Hungarian COVID-19-related words with negative connotations, such as *COVIDszopás* 'annoyance/unpleasant situation due to COVID' (*szopás* means 'sucking'), *COVID-tálibok* 'COVID-Talibans' (also

karanténtálibok ‘quarantine Talibans’), *COVID-fasizmus* ‘COVID fascism’, *COVID-faszság* or *kovidfaszság* ‘COVID bullshit’, *COVIDgeci* ‘unpleasant situation due to COVID’, etc.

The outbreak of the emergency in Hungary and its *maszk*-related “viral” lexicon

The second COVID-19 wave in Hungary began already in August 2020, and the infections were increasing exponentially. During this second wave, the emphasis was on the importance of wearing face masks (*maszkviselés*). Before the pandemic, the word *maszk* ‘mask’ is present in the corpus in minimal proportions and with a different meaning (‘a covering for the face that hides the person wearing it’), such as in the following example: “The robbers wore masks to hide their identities.” After the outbreak of the pandemic, due to the mandatory wearing of face masks, the use of *maszk* became widespread and its productivity exploded. With respect to *maszk* ‘mask’, *szájmaszk* ‘mouth mask’, and *arcmaszk* ‘face mask’, the word *arcpajzs* ‘face shield’ had no success and it did not spread in the common language (it registered only 478 hits in 2020). Among these words, the most frequent is *maszk* ‘mask’ (inflected forms included) with its 576,731 hits in the corpus. The following paragraphs show the high productivity of the term *maszk* and the frequency of the corresponding mask-related neologisms.

A group of these neologisms denotes different types of masks (*maszktípus* ‘mask type’ with 44 hits) according to (i) the area covered by the mask, (ii) its functions, (iii) the materials it consists of, etc.:

(i) *szájmaszk* ‘mouth mask’ (9,774);[9] *arcmaszk* ‘face mask’ (3,110); *orr-szájmaszk ‘nose-mouth mask’* (20); *orrmaszk* ‘nose mask’ (5).
(ii) *védőmaszk* ‘protective mask’ (3,605); *oxigénmaszk* ‘oxygen mask’ (152); *sebészmaszk* ‘surgical mask’ (100); *légzőmaszk* ‘respirator, breath mask’ (19); *vírusmaszk* ‘virus mask’ (10); *műtősmaszk* ‘surgical mask’ (5).
(iii) *textilmaszk* ‘textile mask’ (413); *szövetmaszk* ‘mask made with fabric’ (142); *FFP-maszk* ‘FFP mask’ (23); *pamutmaszk* ‘cotton mask’ (19); *pleximaszk* ‘plexiglass mask’ (18); *vászonmaszk* ‘canvas mask’ (17); *FFP3-maszk* ‘FFP3 mask’ (5).
(iv) *csodamaszk* ‘miraculous mask’ (5).

A second group concerns the act of wearing the mask over the nose, mouth, and chin. The noun *maszkviselés* ‘wearing a mask’ (9,913) refers to the act of wearing a mask such as in *A maszkviselés kötelező marad kültéren is*. ‘Wearing (face) masks remains mandatory also outdoors’. In the corpus, there are different synonyms and

9 The numbers between brackets indicate the number of hits registered in 2020.

variants: *maszkhasználat* 'usage of masks' (1,805) and *szájmaszkhasználat* 'usage of mouth masks' (7); *maszkhordás* 'wearing of masks' (338); *maszkviselet* 'wearing of masks' (286) and *szájmaszkviselet* 'wearing of mouth masks' (5); *szájmaszkviselés* 'wearing of mouth masks' (71); *arcmaszkviselés* 'wearing of face masks' (8). To these abstract nouns we can add the derivational suffix *-i* to create adjectives: *maszkviselési* (1246) and *maszkviselési-* (4) 'mask wearing'. An example is *maszkviselési szabályok* 'mask wearing rules'. Other variants are *maszkhasználati* 'mask usage' (61), such as in *maszkhasználati szabályok* 'mask usage rules' or *maszkhordási* 'mask wearing' (42), cf. *maszkhordási fegyelem* 'mask wearing discipline'; *szájmaszkviselési* 'mouth mask wearing' (11). More complex neologisms are the abstract noun *maszkviselés-ellenesség* 'anti-mask wearing' (2) and the adjective *maszkviseléses* 'mask wearing' (2), such as in *maszkviseléses élet* 'mask wearing life'.

A person wearing a mask is *maszkos* 'masked' (1,272); *szájmaszkos* 'masked with mouth mask' (137); *védőmaszkos* 'masked with protective mask' (35); *arcmaszkos* 'face masked' (22); *maszkos-kesztyűs* 'masked and gloved' (12).

The Hungarian word *maszkviselő* (34, present participle) can be used as an adjective or as a noun, cf. *Jómagam maszkviselő állampolgár vagyok* 'I am a mask wearing citizen' or *Tudatos maszkviselő vagyok* 'I am a conscious mask wearer'.

The person who is not wearing a mask is *maszktalan* (*adj*) 'without a mask' (20), where *-talan* is a privative suffix or *maszknemviselő* 'person not wearing a mask' (4). It is possible to add to the adjective *maszktalan* another derivational suffix to create the corresponding abstract noun *maszktalanság* 'the condition of wearing no mask' (2). The two compounds *maszknélküliség* (4) and *maszk-nem-viselés* 'non-mask-wearing' (1) have the same meaning. The last word can function also as a base for another adjective, *maszknemviselési* (*adj*) 'not wearing masks' (4), cf. *maszknemviselési vita* 'debate around not wearing masks'.

An adverbial derivational suffix may be added to the adjective *maszkos* 'masked' as well: *maszkosan* 'in mask; wearing a mask' (21). E.g.: *Az viszont határozottan jó, hogy a diákok maszkosan nem tudnak cigarettázni!* 'On the other hand, it is definitely good that students cannot smoke when wearing masks!'

In Hungarian, a person (or a group) that does not agree with wearing masks and spreads and encourages opinions against it is defined as being *maszkellenes* (180, *n*, *adj*); *maszkellenző* (3) or *maszktagadó* (92) 'anti-mask', e.g. *Magyarországon is vannak komoly maszkellenes csoportok* 'There are also serious anti-mask groups in Hungary'; *Nem véletlen a rengeteg vírusszkeptikus és maszkellenes, ám ezek nem a megfelelő reakciók egy ilyen válság idején.* 'It's no coincidence that there are plenty of viral skeptics and anti-masks, but these are not the right reactions in a time of such a crisis.' Among the neologisms, there is also *maszkszkeptikus* 'mask skeptical' (7) and *maszkhasználat-ellenes* 'that does not agree with using masks' (6). A particularly complex neologism is *maszkellenes-vírustagadó-konteós* 'anti-mask, virus denier, conspiracy theorist' (2).

With its 49 occurrences, the adjective *maszkmentes* ‘mask-free’ is also rather frequent in the corpus, e.g. *A vírustagadók egy része péntekre maszkmentes napot hirdetett.* ‘Some of the virus deniers declared a mask-free day for Friday.’ The corresponding abstract noun *maszkmentesség* ‘the state of being mask-free’ (5) is rather rare.

On the contrary, a person who agrees with wearing masks is *maszkrajongó* ‘fan of masks’ (6); *maszkpárti* ‘pro-mask’ (21) or *maszkhívő* (3), e.g. *Heves összetűzések voltak országszerte a maszkpárti és maszkellenes tábor között, gyakran az üzletek előtt.* ‘There were fierce clashes across the country between the pro-mask and anti-mask groups, often in front of shops.’

In the corpus we can also find the abstract nouns *maszktagadás* ‘negation of masks’ (15) and *maszkellenesség* ‘the condition of being anti-mask’ (16), and also *maszkvita* ‘mask debate’ (16) and *maszkháború* ‘mask war’ (8); the adjectives *maszk-elutasítási* ‘mask refusing’ (e.g. *a maszk-elutasítási hajlandóság az életkorral csökken* ‘the propensity for mask refusal decreases with age’); *maszkelutasító* ‘mask refusing’; *maszktagadós* (*adj*) ‘mask-denier’; *maszktalanítva* ‘unmasked; a person whose mask was removed’.

Maszk- ‘mask’ is present 130 times in wider expressions (with omissions), e.g. *Emellett a hétvégén újra kötelező a maszk- és kesztyűviselés.* ‘In addition, wearing a mask and gloves is mandatory again over the weekend.’

The compound noun *maszkgyártás* ‘production of mask’ is present in the 2020 Timestamped corpus 123 times, while the related *maszkgyártó* ‘producer of masks’ is recorded 120 times and *maszkgyár* ‘mask factory’ 7 times. It is also possible to find hits for *maszkgyáros* ‘mask manufacturer’ (4); *maszkgyáras* ‘manufacturer of masks’ (2), and *maszkgyár-látogatás* ‘mask factory visit’ (2). In the corpus, there are also a few *hapax legomenon* (with only 1 hit): *maszkgyári* (*adj*) ‘mask factory’ (e.g. *Trump még a maszkgyári programjára sem vett fel maszkot* ‘Trump did not even wear a mask during his mask factory visiting program’); *maszkgyártási* ‘mask manufacturing’ or *maszk-gyártó-gép* ‘mask making machine’.

In addition, in the corpus, there are neologisms concerning:

(i) the necessity to sew masks [*maszkvarrás* ‘mask sewing’ (95); *maszkvarró* ‘sewer of masks’ (32); *maszkvarrógép* ‘mask sewing machine’ (5); *szájmaszkkészítő* ‘mouth mask maker’ (5);], and

(ii) the initial worldwide mask shortage and the necessity to provide masks to the population [*maszkhiány* ‘lack of mask’ (70); *maszkigény* ‘mask demand’ (3); *maszk-osztás* ‘distribution of masks’ (43); *maszkkészlet* ‘mask stock’ (18); *maszkbe-szerzés* ‘purchase of masks’ (17); *maszkvásárlás* ‘mask purchase’ (15), *maszkpiac* ‘mask market’ (9); *maszkadomány* ‘mask donation’ (5); *maszkoszto-gatás* ‘mask distribution’ (5); *maszkdiplomácia* ‘mask diplomacy’ (59); *maszk-kalózkodás* ‘mask piracy’ (1); *maszkügy*, *maszk-ügy* ‘mask affair’ (18); *maszkszállítmány* ‘mask shipment’ (35); *maszkkiállítás* ‘mask exposition’ (5); *maszk-eladási* (*adj*) ‘mask sale’ (e.g. *Múlt hét végén drasztikusan megugrottak*

a hazai gyógyszertárakban az orvosi maszk-eladási számok 'Last weekend, the numbers referring to the surgical mask sales in Hungarian pharmacies increased drastically'); *maszk-szállítás* 'mask delivery' (1); *maszkbiznisz* 'mask business' (6); *maszkeladás* 'mask selling' (3)].

In addition, the corpus contains 58 mask-related rare neologism and 53 *hapax legomena*. All these creations are included in TCD, even if this kind of data is usually left out of dictionaries. These *hapax* play an important role in the assessment of productivity and creativity of the Hungarian language. It is well known that the lifespan of a neologism is, from the moment of its first appearance, uncertain and difficult to predict: some of the neologisms seem destined to last, while others are not. In the long term, every forecast turns out to be uncertain and there is the risk of excluding neologisms destined for success. So, considering the neologisms containing *korona-* 'corona-', *COVID-*, *járvány-* 'epidemic', *vírus-* 'virus', *maszk-* 'mask', *karantén-* 'quarantine', and *oltás* 'vaccine' as constituents, I have decided to systematically collect all the new words encountered, without taking into account their actual use and the degree of their diffusion. In fact, there is a risk of including too many entries in TCD, but there is also the advantage of identifying with greater precision the paths of neological activity. Later in time, there will be the possibility to understand the reasons for the success of some of these new words and to discuss the predictable failure of many occasional and ephemeral creations. In any case, I consider it useful to record all of these neologisms, *hapax* included, even if I am aware that their neological status is objectively less strong and sustainable.

In total, in the corpus, there are more than 210 mask-related neologisms and only 9 are formed by 'simple' derivation, 1 is the lemma *maszk*, and the remaining parts are compounds and may be the result of multiple derivation. The neologisms *maszkné* 'maskne' and *maszkitisz* 'maskitis' are blend words (*maszk* + *akné*; *maszk* + *dermatitisz*) and they entered into the Hungarian language from English. *Aranymaszk* 'gold mask' is an occasionalism. It refers to the story of the businessman Shankar Kurhade who bought a customized gold mask. *Okosmaszk* 'smart mask' is formed following the examples of *okostelefon* 'smart phone', *okoseszköz* 'smart device', *okosszemüveg* 'smart glasses', *okosóra* 'smartwatch', etc.

All of the words are accompanied by useful child elements providing indications concerning frequency, such as 'frequent', 'rare', 'hapax', and frequency of use in time, such as a particular date (the point in time when a word started to be used), first, second, etc. wave or other information related to the trends of the words treated (unusual increase or decrease in use). For each entry, at least one translational equivalent will be provided in Italian and English. In Figure 1, the entry *maszkvita* 'mask debate' illustrates the microstructure of TCD.

---

maszkvita; *n*
Definition pro és kontra a szájmaszkokról
Word formation process compounding
Frequency rare
Trend peaks in March and September/October 2020
Temporal use from March 2020
Register formal
Connotative effect
Subject field
Style literal
Etymology
References
— ♦ *Van olyan magyar település egyébként, ahol a polgármester döntötte el a* ***maszkvitát****: a koronavírus miatt szerdától csak maszkban mehetnek az emberek boltba és más üzletekbe a Fejér megyei Velencén.*
♦ *Emiatt húzódott sokáig a* ***maszkvita*** *is, de most már szinte mindenki azt javasolja, elővigyázatosságból hordjunk maszkot ott, ahol másokkal találkozunk.*
Note *maszk + vita "debate, discussion"*
It. dibattito sulle mascherine; *noun mwe*
Etymology
References
♦ *Il* ***dibattito sulle mascherine*** *è strettamente legato a un'altra questione che ha suscitato forti divisioni: in che modo il virus si sposta nell'aria e diffonde l'infezione?*
♦ *La variante Delta riaccende il* ***dibattito sulle mascherine*** *negli Usa*
Note
Eng. mask debate; *noun mwe*
Etymology
References
♦ *Delta variant reignites US* ***mask debate***
♦ ***Mask debate*** *From School Boards to Courtrooms.*
Note

Figure 1: Entry *maszkvita* 'mask debate'.

If, due to translational difficulties, no equivalent can be given, a descriptive/explanatory equivalent is added (cf. Figure 2), e.g. the word *nyunyóka* 'comfort object' has been in use since May 13, 2020, with peaks in May when it was used in the Chief Medical Officer Cecília Müller's discourse and in December 2020 when it was named the word of the year.

nyunyóka; *n*
Definition A nyunyóka is a lovey and it can be anything that is safe for a baby or toddler to have at sleep time. It is a sort of comfort or transitional item – blanket or stuffed animal or other comfort object of affection that a baby or toddler brings to bed, and that provides comfort and soothing.
Orthographic variants -
Word formation process derivation
Frequency Before May 13, 2020, the term nyunyóka was uncommon and was used only in baby talk.
Temporal use since May 13, 2020, as a result of the massive media impact of Chief Medical Officer Cecília Müller's discourse during the press conference of the Operational Force, concerning personal hygiene habits to teach kids and the necessity to wash comfort objects frequently, this neologism entered the common language and became widely known and used.
Trend Increasing. The results of the search query using Google now lists 46,500 pages (as of December 17, 2021).
Register baby talk > common language
Connotative effect term of endearment
Subject field
Style
Etymology Mind a nyanya, mind a nyunyó föltehetőleg dajka-, gyermeknyelvi, hangulatfestő és talán hangutánzó szó. Elképzelhető a nyúl, nyuszi szavak becézéséből fakad. Ezt erősíti, hogy a neten nyuszi-nyunyi szundikendő is rendelhető. https://e-nyelvmagazin.hu/2021/01/10/nyunyoka-nyunyo/
References Veszelszki 2020: 54–55.
— ♦ *Kerülni kell, hogy a gyerek otthonról játékot hozzon el: pelenka, **„nyunyóka”**, játék maradjon otthon, ám azokat otthon is tisztítsuk rendszeresen.*
Note *Nyunyóka: Kisgyermek alvó játékszere, leginkább plüssfigura. Másnéven: alvóka, rongyi. Hangtani rokona a nyanya, a nagymama kedveskedő, az öregasszony gúnyos megnevezése.*
It. doudou; *n*
Etymology
References
Note oggetto transizionale
Eng. lovey; *n*
Etymology
References
Note comfort object, transitional object

Figure 2: Entry *nyunyóka* ‘comfort object’.

5 Conclusions

The high number of coroneologisms draw attention to the creativity and vitality of the Hungarian language in times of crisis, and the corpus analyses performed for the *Trilingual (HU, IT, EN) COVID-19 Dictionary* provide a clearer picture of the change in

the vocabulary during the COVID-19 pandemic and of the role and function of word formation processes that contributed to the creation of these neologisms. The analyses would suggest that the most frequently occurring word formation processes of the Hungarian neologisms related to the pandemic are compounding and derivation, but syntagms, blending, and semantic extension (changes in lexical meaning) are also used. [Furthermore, in Hungarian, new words may be productively created also by means of conversion, backformation, reduplication, clipping, loan words, and loan formations (e.g. calques), metaphor . . .] At the end of the pandemic, the analyses will also reveal to what extent the Hungarian language borrows coroneologisms from other languages.

Time passes, but the impact of the COVID-19 pandemic on the Hungarian language is still strong in 2021. It is true that many terms had their peaks during the first months of the crisis in 2020, but each phase and wave produce new topics, and terms, and generate considerable frequency increases in the use of certain forms. Therefore, the corpus-based dictionary will be a valuable tool to explore and analyze the corona-lexicon in the Hungarian press and common language during this global emergency thanks to the microstructure of the entries that, where possible, includes information on frequency, trends, and temporal use. The entries contain a morphological analysis too, so that TCD provides data that will help to analyze the trends and patterns in the formation of new words and in their frequency of use in Hungarian.

Overall, this dictionary is useful for linguists and translators (e.g. suggesting more accurate translation equivalents for translating the coroneologisms from Hungarian to English or Italian) or for scholars in the digital humanities.

Bibliography

Balázs, Géza (2020): A koronavírusról szóló beszéd (nyelv és folklór). In Kovács, László (ed.): *Globális kihívás – lokális válaszok. A koronavírus (Covid19) gazdasági és társadalmi összefüggései és hatásai*. Szombathely: Savaria University Press, 229–240.

Bušta, Jan, et al. (2017): JSI Newsfeed Corpus. In: *The 9th International Corpus Linguistics Conference*, *July* 25*–28, 2017, Birmingham, GB: extended abstracts*. Birmingham: University of Birmingham, [https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper382.pdf; last access July 29, 2021].

Coroneologisms are going viral. In: *Economic Times*. 9 April 2020. [https://economictimes.indiatimes.com/blogs/et-editorials/coroneologisms-are-going-viral/ last access June 30, 2020].

Dajka, Balázs (2020): Lecsapott a koronavírus az amúgy is nehéz időket élő kézicsapatra. In: *24.hu*. October 20, 2020. [https://24.hu/sport/2020/10/20/kezilabda-koronavirus-karanten-siofok-mtk/; last access July 29, 2021].

Domonkosi, Ágnes/Ludányi Zsófia (2020): Társas távolságtartás és nyelvi közeledés – e-mailezési gyakorlatok a koronavírus idején. In: Kovács, László (ed.): *Globális kihívás – lokális válaszok. A koronavírus (Covid19) gazdasági és társadalmi összefüggései és hatásai*. Szombathely: Savaria University Press, 241–260.

Elszabadult a koronavírus-járvány a világon: egy nap alatt 400 ezer fertőzött. In: *Portfolio.hu*. 19 October 2020. [https://www.portfolio.hu/gazdasag/20201019/elszabadult-a-koronavirus-jarvany-a-vilagon-egy-nap-alatt-400-ezer-fertozott-453448#; last access July 29, 2021].
Hát ezért is olyan veszélyes, hogy berobbant Magyarországon a koronavírus. In: *Portfolio.hu*. 20 October 2020 [https://www.portfolio.hu/gazdasag/20201020/hat-ezert-is-olyan-veszelyes-hogy-berobbant-magyarorszagon-a-koronavirus-453454; last access July 29, 2021].
Istók, Béla/Lőrincz, Gábor (2020): A virolingvisztika részterületei. In Simon, Szabolcs (ed.): *12th International Conference of J. Selye University. Language and Literacy Section. Conference Proceedings*. J. Selye University. Komárno. 83–92. [DOI: 10.36007/3761.2020.83; last access July 29, 2021].
Kegyes, Erika/Lanzmaier-Ugri, Katharina (2020): A többnyelvű tájékoztatás kihívásai, stratégiái és eredményei a koronavírus idején Ausztriában. In Kovács, László (ed.): *Globális kihívás – lokális válaszok. A koronavírus (Covid19) gazdasági és társadalmi összefüggései és hatásai*. Szombathely: Savaria University Press, 261–289.
Kilgarriff, Adam (2001): Web as corpus. In Rayson, Paul et al. (eds.): *Proceedings of the Corpus Linguistics 2001 Conference, Lancaster University, March 29–April 2, 2001*. Lancaster: UCREL, 342–344.
Kilgarriff, Adam/Greffenstette, Gregory (2003): Introduction to the Special Issue on the Web as Corpus. In: *Computational Linguistics* 3(29), 333–347. [DOI: 10.1162/089120103322711569; last access July 29, 2021].
Kovács, László (ed.) (2020): *Globális kihívás – lokális válaszok. A koronavírus (Covid19) gazdasági és társadalmi összefüggései és hatásai*. Szombathely: Savaria University Press.
Lénárt, István (2020): Karanténykedünk: a karanténdepitől a karanténszexig – avagy hogyan árulkodnak tevékenységeinkről a Koronavírus-járvány alatt született szóösszetételek. In Kovács, László (ed.): *Globális kihívás – lokális válaszok. A koronavírus (Covid19) gazdasági és társadalmi összefüggései és hatásai*. Szombathely: Savaria University Press, 291–302.
Ludányi, Zsófia (2020): Helyesírási kérdések pandémia idején. In: *Amega: asztma és allergia színes tájékoztató magazin* 4(27), 33–35.
Ludányi, Zsófia (2021): Eloltották az oltóanyagot. In: *Amega: asztma és allergia színes tájékoztató magazine* 2(28), 37–39.
Ludányi, Zsófia (2021): Oltakozás és pándémia – avagy a virolingvisztika legújabb kérdései. In: *AMEGA: asztma és allergia színes tájékoztató magazin* 1(28), 43–45.
Měchura, Michal Boleslav (2017): Introducing Lexonomy: an open-source dictionary writing and publishing system. In: Kosem, Iztok et al. (eds): *Electronic lexicography in the 21st century: Lexicography from Scratch. Proceedings of the eLex 2017 conference, September 19–21, 2017*. Leiden: Lexical Computing, 662–679.
Papp, Judit (2021): il nostro lessico è diventato "virale". Il vocabolario dell'emergenza sanitaria, economica e sociale ai tempi della pandemia di COVID-19. In: *Studia Universitatis Babes-Bolyai – Philologia* 1(2021), 325–344.
Roig-Marín, Amanda (2020): English-based coroneologisms: A Short Survey of Our Covid-19-Related Vocabulary. In: *English Today*, 1–3. DOI: 10.1017/s0266078420000255.
Trampus, Mitja/Novak, Blaz (2012): The internals of an aggregated web news feed. In: *Proceedings of the Fifteenth International Information Science Conference IS SiKDD*. October 8–12, 2012, [Ljubljana, Slovenia], 431–434.
Tyrrell, David Arthur John/Michael Fielder (2002): *Cold Wars: The Fight Against the Common Cold*. Oxford: Oxford University Press.
Veszelszki, Ágnes (2020): *Karanténszótár*. Budapest: Interkulturális Kft.
Virology. Coronaviruses. In: *Nature* 220 (November 16, 1968): 650. [https://doi.org/10.1038/220650b0; last access 29 July 2021].

Milica Mihaljević, Lana Hudeček, Kristian Lewis

Coronavirus-related neologisms: A challenge for Croatian standardology and lexicography

1 Introduction

Languages have always evolved to reflect societal changes, but in 2020–2021 their evolution could be seen in real-time. The appearance of Coronavirus[1] has led to an abundance of new words and phrases, both in Croatian and other languages.[2] As stated by Lawson,“[t]his new vocabulary helps us make sense of the changes that have suddenly become part of our everyday lives.” (Lawson 2020).

Recent epidemics have given rise to the appearance of new words and phrases which were not necessarily coined for the current COVID-19[3] pandemic. Still, those words and phrases have gained a far wider usage since 2020, e.g., “the term *infodemic* was coined in 2003 for the SARS epidemic, but has also been used to describe the current proliferation of news around coronavirus.” (OED Blog 2020).

Moreover, many terms used mostly by medical experts have entered the general language (e.g. *Coronavirus*, *epidemiology*, *asymptomatic*). From spring 2020, laypersons have become familiar with terms that have been around for years but have not been used in the general language, some even dating from the 19th century but having achieved new and much wider usage. For example, “*self-isolation* (recorded from 1834) and *self-isolating* (recorded from 1841) are now used to describe self-imposed isolation to prevent catching or transmitting an infectious disease, where in the 1800s these terms were more often applied to countries which chose to detach themselves politically and economically from the rest of the world.” (OED Blog 2020).

1 In English, this term is spelled *Coronavirus* and *coronavirus*. In this paper, we use *Coronavirus*.
2 E.g., for Macedonian cf. Janusheva (2020), for Russian cf. Karachina (2020).
3 In English, this term is spelled *Covid-19* and *COVID-19*. In this paper, we use *COVID-19*.

Note: This paper is written within the research project *Croatian Web Dictionary – Mrežnik*.

Milica Mihaljević, Institute of Croatian Language and Linguistics, Ulica Republike Austrije 16, Zagreb, Croatia, e-mail: mmihalj@ihjj.hr
Lana Hudeček, Institute of Croatian Language and Linguistics, Ulica Republike Austrije 16, Zagreb, Croatia, e-mail: lhudecek@ihjj.hr
Kristian Lewis, Institute of Croatian Language and Linguistics, Ulica Republike Austrije 16, Zagreb, Croatia, e-mail: klewis@ihjj.hr

https://doi.org/10.1515/9783110798081-009

The media and social media have an important role in the appearance and spreading of new Coronavirus-related words, phrases, and meanings and old terms used in a new context. The appearance of Coronavirus was followed by Coronavirus jokes,[4] memes, and puns e.g., *Ljubav u doba kolere* > *Ljubav u doba korone* > *Život u doba korone* 'love in the time of cholera' > 'love in the time of corona' > 'life in the time of corona'.[5] Even some Coronavirus-related nicknames appeared in Croatian media and social media. For example, *Apaurin*[6] *Hrvatske* 'Apaurin of Croatia' was a nickname for the Croatian epidemiologist Alemka Markotić meaning that she can calm down Croatia.[7] *Toni Cjepinski* 'Toni the vaxxer' is a nickname for Croatian singer Toni Cetinski who is well-known for his opposition to vaccination.[8]

The appearance of new words and phrases, new meanings of existing words, and the shift of words and phrases from (mostly medical) scientific terminology to the general language has been studied by linguists from different perspectives – cognitive linguistics, lexicology, ethnolinguistics, phraseology, corpus linguistics, etc. "Even a superficial glance at public discourse on the pandemic reveals that it is saturated with metaphors (we talk about epidemic epicentre, epidemic focal point, the wave of the epidemic, modern plague and flaming epidemic), and especially with war metaphors (words like headquarters, first line of defence, invisible enemy and the war against the virus)." (Štrkalj Despot/Ostroški Anić 2021: 3).

2 Methodology

This paper focuses on standardological and lexicographical aspects of Coronavirus-related neologisms in Croatian. The presented results are based on corpus analysis. The initial corpus for this analysis consists of terms collected for the *Glossary of Coronavirus*. This corpus has been supplemented by terms we collected on the Internet and from the media. The General Croatian corpora: *Croatian Web Corpus* – hrWaC (cf. Ljubešić/Klubička 2016) and *Croatian Language Repository* (cf. Brozović Rončević/Ćavar 2008: 173–186) were also used, but since they do not include neologisms that entered the language after 2013, they could be used only to check terms in the language before that time. From October 2021, a specialized Corona corpus compiled by Štrkalj Despot and Ostroški Anić (2021) became publicly available on

4 Coronavirus jokes in Croatian were analyzed by Miloš (2020).

5 Štrkalj Despot (2020) lists such puns in Croatian and mentions that the phrase *u doba korone* 'in the time of Corona' has more than 4 million hits on Google.

6 *Apaurin* is the name of an anxiolytic.

7 However, as the pandemic continued, the nickname disappeared as quickly as it appeared.

8 In Croatian *cijepiti se* means 'to vaccinate'.

request.[9] The data from these corpora are analyzed by Sketch Engine (cf. Kilgarriff *et al.* 2004: 105–116), a corpus query system loaded with the corpora, enabling the display of lexeme context through concordances and (differential) word sketches and the extraction of keywords (terms) and N-grams. The most common collocations are sorted into syntactic categories. For English equivalents, in addition to the sources found on the Internet, *enTenTen2020* corpus was consulted.

In the second part of the paper, we analyze and compare the presentation of Coronavirus terminology in the descriptive *Glossary of Coronavirus* and the normative *Croatian Web Dictionary – Mrežnik*.

3 Loanwords, loan translations or Croatian neologisms

The COVID-19 pandemic caused the appearance of many Coronavirus-related neologisms in many languages. Due to the speed and intensity of this infection, many Croatian words have been directly borrowed from English (cf. Štrkalj Despot 2020: 2).

Many Coronavirus-related neologisms in Croatian are either loanwords or loan translations from English. In some cases, it is difficult to determine whether a particular term was independently formed in Croatian or is a loan translation from English. Table 1 shows some Coronavirus-related loanwords which have entered Croatian with no or little adaptation to the Croatian language system.[10]

English Coronavirus-related neologisms are often compounds/multi-word units,[11] blends, and abbreviations. English has many compounds with the first element

9 "First, we compiled a specialized corpus of Croatian media texts (referred to here as the *Korona* corpus) using Sketch Engine corpus compilation tools. The corpus consists of manually selected texts dated January 29 to December 23, 2020, all closely related to the coronavirus Covid-19 pandemic topics. For free access to the corpus, please write to the authors." (cf. Štrkalj Despot/Ostroški Anić 2021: 180). Mihaela Matešić announced the compilation of another Croatian Corona-related corpus in her presentations *Analiza stavova u društvenim medijima u istraživanju krizne komunikacije u doba pandemije* (Analysis of attitudes in social media in the research of crisis communication during the pandemic) in Rijeka, June 23, 2021 (5. simpozij SCIMETH) https://cji.uniri.hr/scimeth-2021/) and *Istraživanje krizne komunikacije u digitalnom okruženju u doba pandemije* (Exploring crisis communication in the digital environment during the pandemic) (with Slobodan Beliga), Osijek, September, 9–11 2021 http://www.hdpl.hdpl.hr). On the same symposium Mateja Šporčić, Stjepan Lacković, and Marina Baralić in the presentation *Big data resursi u istraživanju metafora* (Big data resources in the study of metaphors) announced the compilation of yet another Corona corpus. However, these corpora are in November 2021 not yet publicly available.

10 According to Croatian orthographic rules, unadapted loanwords are spelled in italic.

11 It is often difficult to differentiate between compounds and multi-word units in English as the spelling of these terms varies – one or two words.

Corona. The same model is very productive in Croatian Coronavirus-related neologisms. In Table 2, some compounds with the first element *Corona* are given in English and Croatian.

Table 1: English loanwords in Croatian.

English	Croatian
drive-in	*drive-in*
fake news	*fake news*
infodemia	infodemija
targeted	targetiran
lockdown	*lock down*
Corona free	korona *free*

Table 2: Compounds with the element *Corona* in English and Croatian.

English[12]	Definition	Croatian
Corona babies/ Corona boomers	the generation born after December 2020 and who will become teenagers in 2033–2034; the terms *coronals* and *quaranteens*[13] have the same meaning	koronijalci
Corona bed	hospital bed with a ventilator for Covid patients	koronakrevet
Corona crew	people one chooses to live with during the quarantine; the term *quaranteam* has the same meaning	–
Corona crisis	crisis caused by the COVID-19 pandemic	koronakriza
Corona cushion	fat acquired during lockdown due to COVID-19	koronašpek
Corona haircut	haircut performed by oneself or another unqualified person	koronafriz
Corona humor	jokes and memes which appeared on (social) media during the COVID-19 pandemic	koronahumor
Corona partner/ boyfriend/girlfriend	a partner during the lockdown period that implies brevity of the relationship	–
Corona rules	rules one has to obey during the COVID-19 pandemic	koronapravila

12 We have consulted Alyeksyeyeva/Chaiuk/Galitska (2020) and Khalfan/Batool/Shehzad (2020) for definitions of these terms.
13 The term *quaranteens* sometimes has another meaning, ‘teenagers during the quarantine’.

Table 2 (continued)

English	Definition	Croatian
Corona speak	neologisms that have entered everyday discourse during the Coronavirus pandemic	–
Corona traffic light	marking of regions by different colors (green, yellow, and red) according to the number of registered newly infected by Coronavirus	koronasemafor

From Table 2, we can see that some of these terms have a Croatian equivalent formed by the same model *Corona + noun* e.g., *Corona bed* is *koronakrevet*, *Corona haircut* is *koronafriz*, *Corona traffic light* is *koronasemafor*, etc. Some terms that exist in English have not been recorded in Croatian e.g., *Corona boyfriend*, *Corona crew*, *Corona partner*.

Coronavirus-related English neologisms are often formed by blending. As this word-formation type is not typical in Croatian, the equivalent Croatian terms were usually formed differently, as shown in Table 3.

Table 3: Coronavirus-related blends in English with their Croatian equivalents.

English term	Croatian term	Word formation elements and definition
coronacession	recesija uvjetovana koronom ('recession caused by Corona')	*Corona + recession* recession caused by Coronavirus
coronials/ quaranteens	koronijalci	*Corona + Millennials* children conceived or born during the COVID-19 pandemic, reminiscent of the label *millennial* describing those raised around the turn of the millennium (cf. *Words We're Watching: 'Coronial'*)
covidiot	kovidiot	*COVID + idiot* an insulting term for someone who ignores health advice about COVID-19, hoards food unnecessarily, etc. (cf. *Macmillan Dictionary*)
covidivorce	koronarazvod (coronadivorce)	*COVID + divorce* divorce caused by the fact that partners spend too much time together due to COVID-19 lockdown

Table 3 (continued)

English term	Croatian term	Word formation elements and definition
quarantime	vrijeme (provedeno) u karanteni ('time (spent) in a quarantine')	*quarantine + time* the time during COVID-19 when all of the days blend because you are stuck inside due to a worldwide pandemic (cf. *Urban Dictionary*)
quarantini	*quarantini*	*quarantine + Martini* a drink that one drinks during lockdown
she-cession	–	*she + recession* the recession caused by COVID-19, which affected women more than men[14]
twindemic	*twindemija*	*twin + pandemic* flu season paired with COVID-19

If we compare English blends with their Croatian equivalents, we can conclude:

1. Some English blends are used as loanwords in Croatian; some are adapted to Croatian orthography (*covidiot* > *kovidiot*, *coronials* > *koronijalci*), some are only partly adapted (*twindemija*), and some are spelled as in English (*quarantini*);
2. Croatian has compounds (*covidivorce* > *koronarazvod*) as equivalents of English blends;
3. Croatian uses multi-word expressions e.g., *coronacession* > *recesija uvjetovana koronom* ('recession caused by Corona'), *quarantime* > *vrijeme (provedeno) u karanteni*) ('time spent in quarantine') as equivalents of English blends;
4. instead of the possible loanword *karantinci* for *quaranteens* Croatian uses the synonymous term *koronijalci* ('coronials');
5. Croatian does not have any equivalent of *she-cession*.

However, although blending is not common for Croatian these Coronavirus-related blends have been recorded on the Internet: *kupomanija* – *kupovanje* ('shopping') + *manija* ('mania') ('mania for shopping, extensive shopping'), *Zoomor* – *Zoom* + *umor* ('fatigue') ('fatigue caused by Zoom').

In English Coronavirus-related terminology, many abbreviations are used. Abbreviations and multi-word terms from which they are derived are shown in Table 4.

Some of the English abbreviations are also used in Croatian e.g., *COVID*, *SARS-CoV-2*, *PCR*.[15] There are no Coronavirus-related abbreviations from Croatian terms. In Croatian, either the English abbreviation or the full Croatian term is used e.g.,

14 "Of the 20.5 million jobs lost last month, women make up to 55 percent of those now looking for work. Their unemployment rate is about 2 percent higher than that of men." (cf. Andrews 2020).
15 PCR is in Croatian read as in English [ˈpiː-ˈsiː-ˈɑr]. Croatian reading would be [ˈpeː-ˈceː-ˈer].

Table 4: Abbreviations in English and Croatian.

Abbreviation	Multi-word term	Croatian
ARDS	acute respiratory distress syndrome	akutni respiratorni distresni sindrom
CoV	Coronavirus	koronavirus
COVID-19	Coronavirus disease 19 (the year of onset, 2019)	koronavirusna bolest 2019
ECMO	extracorporeal membrane oxygenation	izvantjelesna membranska oksigenacija
MERS	Middle East respiratory syndrome	bliskoistočni respiratorni sindrom
NAAT	nucleic acid amplification test	test koji se temelji na umnažanju nukleinske kiseline
PCR	polymerase chain reaction	polimerazna lančana reakcija
PPE	personal protective equipment	osobna zaštitna oprema
PUI	patient under investigation	osoba pod zdravstvenim nadzorom
SARS-CoV	severe acute respiratory syndrome – Coronavirus	teški akutni respiratorni sindrom – koronavirus
WFH	working from home	rad od kuće

osobna zaštitna oprema (‘personal protective equipment’), *rad od kuće* (‘work from home’) are never abbreviated.

4 Coronavirus-related neologisms in Croatian – descriptive analysis

Croatian Coronavirus-related neologisms can be analyzed according to different criteria:

1. single words vs. multi-word units

Coronavirus-related neologisms can be single words or multi-word units as shown in Table 5.

Single words can be analyzed according to word-formation types and can be divided into compounds, derivatives, and semi-compounds; examples are given in Table 6.

Table 5: Coronavirus-related single words and multi-word units.

Single words		Multi-word units	
Croatian	**English**	**Croatian**	**English**
koronazabava	coronaparty	skupni/kolektivni imunitet	group/collective immunity
samotestiranje	self-testing	drugi/treći/četvrti val	second/third/fourth wave

Table 6: Compounds, derivatives, and semi-compounds.

Compounds		Derivatives		Semi-compounds	
Croatian	**English**	**Croatian**	**English**	**Croatian**	**English**
koronaredar	Corona security guard	koronaš	person who has Coronavirus	delta-soj	delta strain
koronamafija	Corona mafia	antimasker	person who is against wearing masks	delta-varijanta/ delta-inačica	delta variant
kovidpozitivan	positive for Coronavirus	samoizolacija	self-isolation	alfa-soj	alpha strain
prvozaraženi	first person to be infected	superširitelj/ superširiteljica	superspreader (male, female)	alfa-varijanta	alpha variant

2. scientific/technical terms (academic jargon terms) vs. jargon words

Some Coronavirus-related neologisms are scientific terms while others are jargon words; examples of scientific terms entering the general language are shown in Table 7.

Table 7: Scientific terms entering the general language.

Croatian	English
anosmija (or anozmija)	anosmia
antigenski test	antigen test
antivakser/antivakserica	antivaxxer (male, female)
crvena zona	red zone
dezinfekcija	disinfection
epidemiolog/epidemiologinja	epidemiologist (male, female)
fizička distance/fizička udaljenost	social distancing

Table 7 (continued)

Croatian	English
izolacija/samoizolacija	isolation/self-isolation
izravnavanje krivulje	flattening the curve
komorbiditet	comorbidity
probir	screening
respirator	respirator
simptomatski/asimptomatski bolesnik/ bolesnica	symptomatic/asymptomatic patient (male, female)
superširitelj/superšiniteljica	super-spreader (male, female)

Many academic terms (mainly medical terms) have entered the general discourse. Words like *symptomatic/asymptomatic*, *disinfection*, *isolation*, *self-isolation*, and *super-spreader* have become a part of general discourse, along with terms from other disciplines such as *red zone*, *social distancing*, and *flattening the curve*. "The expression *social distancing*, for example, has gone from being a relatively unknown piece of academic jargon to something we hear multiple times a day (although the World Health Organization prefers *physical distancing*). Usage of the phrase *flattening the curve* has increased exponentially. The word *super-spreader* has also spread from mouth to mouth at a dizzying rate." (Mahdawi 2020).

Examples of neologisms in Croatian jargon are shown in Table 8.

Table 8: Jargon neologisms.

Croatian	English
cjepiša	a person who has been vaccinated, a person who is in favor of vaccination
imunitet krda	herd immunity
kašljosram	shame because of coughing
koronafriz	Corona haircut
kliconoša	a person or animal that carries microorganisms that cause disease but has no signs of a disease
koronakašljač	a person who coughs on purpose so people would think he has COVID-19
koronaš/ koronašica	a person (male/female) infected by Coronavirus
koronašpek	Corona fat

Table 8 (continued)

Croatian	English
koronizirati	infect by Corona virus
kovidaš	a person having COVID-19
necjepiša	anti-vaxxer, a person who is against vaccination

If we compare scientific terms entering the general language with jargon terms, we can see that scientific terms are mostly of Latin origin and reflect the English term, while jargon terms are mostly derivatives formed by suffixation from Croatian elements. However, there are many jargon terms with the first element *Corona*, which closely mirror the English model.

3. new words vs. words getting a new specialized meaning often through metaphorisation or specialization of meaning
The Coronavirus pandemic caused the appearance of many new terms and additional meanings of the existing words; examples are given in Table 9.

Table 9: New words and new meanings.

New words		New meanings	
Croatian	**English**	**Croatian**	**English**
koronakriza	Corona crisis	balon, balončić, mjehurić	bubble
kovidaš	a person having COVID-19	zatvaranje	lockdown

4. terms denoting disease and diagnosis, terms denoting human reactions, and new way of life
Terms can be roughly divided according to their meaning into these groups: 1. the disease, diagnosis, and disease-related terms; 2. human reactions to the disease and behavioral patterns; 3. new work and school life; examples are given in Table 10.

5 Questions from native speakers

From the beginning of the pandemic in February 2020, the Institute of Croatian Language and Linguistics received many questions from the media but also the general public asking for language advice. Soon it became apparent that Coronavirus-related terminology presented many problems for Croatian speakers. For example, many

orthographic variants of the same term were simultaneously used in the media, sometimes even in the same text e.g., *COVID bolnica*, *covid bolnica*, *Covid bolnica*, *COVID-bolnica*, *covid-bolnica* ('COVID hospital'). Some of these problems occur also in English but when orthographic variants are combined with lexical synonyms many terms denoting the same concept were recorded e.g., the terms *COVID potvrda*, *COVID-potvrda*, *covid potvrda*, *covid-potvrda*, *COVID certifikat*, *COVID-certifikat*, *covid certifikat*, *covid-certifikat*, *COVID putovnica*, *COVID-putovnica*, *covid putovnica*, *covid-putovnica* are all equivalents of the English term *COVID certificate*.

Table 10: Terms divided according to meaning.

Disease, diagnosis, and disease-related terms		**Human reactions, prevention**		**New life**	
Croatian	**English**	**Croatian**	**English**	**Croatian**	**English**
koronavirus	Coronavirus	laktovanje/ laktarenje	elbow bump	fizička/socijalna udaljenost/distanca	social distance
nulti pacijent	patient zero	samoizolacija	self-isolation	hibridna nastava	hybrid learning / blended learning
polumaska	half-mask	potvrda o cijepljenju	certificate of vaccination	rad od kuće	work from home

Most of the questions could roughly be divided into two groups: 1. the correct spelling of certain terms, 2. the correct Croatian term for an English term. Table 11 shows some characteristic questions from native speakers divided according to orthographic, grammatical, and lexical levels.

Table 11: Questions from native Croatian speakers.

Level	**Questions**	**Answers**
Orthographic level	What is the correct spelling of words derived from the word *cijepiti* ('vaccinate')?	*docijepiti/docjepljivati* ('revaccinate'), *procijepljenost* ('vaccination rate'), *cijepljenik* ('vaccinee'), *cjepivo* ('vaccine'), *cjepitelj* ('vaccination administrator'), *cjepilište* ('vaccination point') (cf. Lewis 2021: 28–33)
	What is the correct spelling of the names of some vaccines?	*Pfizer-BioNTech*, *AstraZeneca-Oxford*, *Johnson & Johnson*, *Sputnjik* (cf. Lewis 2021: 28–33)

Table 11 (continued)

Level	Questions	Answers
	What is the correct spelling of terms with the element *COVID*?	*COVID incidencija* ('COVID incidence'), *COVID kredit* ('COVID loan'), *COVID krevet* ('COVID bed'), *COVID negativan* ('COVID negative'), *COVID pacijent* ('COVID patient'), *COVID planinar* ('COVID climber'), *COVID pozitivan* ('COVID positive'), *COVID redar* ('COVID monitor'), *COVID situacija* ('COVID situation'), post-COVID
	What is the correct spelling of terms with the element *korona-* ('Corona-')?	The element *korona-* should be spelled together with the next element *koronabolnica* ('Corona hospital'), *koronafobija* ('Corona phobia'), *koronahumor* ('Corona humor'), *koronakriza* ('Corona crisis'), etc.
	What is the correct spelling of terms *alfa(-)soj* ('alpha variant'), *delta(-) varijanta*, *delta(-)inačica* ('delta variant'), *delta(-)plus(-)soj* ('delta plus variant')?	The correct spelling is *alfa-soj*, *delta-varijanta*, *delta-inačica*, and *delta plus soj*.
Morphological level	How should we decline *COVID-19*?	genitive: *COVID-a 19*; dative: *COVID-u 19*
	How should we decline the names of vaccines?	genitive: *Pfizer-BioNTecha*, *AstraZeneca-Oxforda*, *Johnsona & Johnsona*; dative: *Pfizer-BioNTechu*, *AstraZeneca-Oxfordu*, *Johnsonu & Johnsonu* (cf. Lewis 2021: 28–33)
Word-formation	What is the correct adjective derived from the noun *epidemija* ('epidemic') and *epidemiologija* ('epidemiology')?	The adjective *epidemiološki* ('relating to epidemiology') is derived from *epidemiologija* ('epidemiology'), and *epidemijski* ('relating to the epidemic') is derived from *epidemija* ('epidemic'). The measures are *protuepidemijske* ('anti-epidemic').
	Is it better to say *antigen-testovi* or *antigenski testovi* ('antigene tests')?	The Croatian term should be *antigenski testovi*.
	Is it correct to say *eksponencijalni* or *eksponencijski rast* ('exponential growth')?	*eksponencijski rast*

Table 11 (continued)

Level	Questions	Answers
Syntactic level	Is it correct to say *NAAT test?*	As *T* in *NAAT* means 'test', *NAAT test* is a pleonasm and only *NAAT* should be used.
	What is the correct gender of the adjective in the term *glasnički*/*glasnička mRNK* ('messenger mRNA')?	As acronyms except those ending in *a* are of masculine gender the correct term is *glasnički mRNK*.
Lexical level	Is there a Croatian term for *socijalna distance* ('social distance'), *lokalna transmisija* ('local transmission'), *imunitet krda* ('herd immunity'), *drive-in testiranje* ('drive-in testing')?	Croatian terms are *fizička udaljenost*, *lokalni prijenos*, *kolektivni imunitet provozno testiranje*.
	Are the words *karantena* ('quarantine') and *izolacija* ('isolation') synonyms?[16]	*Izolacija* ('isolation') is the separation of a person or a group of people from the rest of the population. *Samoizolacija* ('self-isolation') is a form of self-imposed isolation. *Karantena* ('quarantine') is a legally imposed restriction of movement of people exposed to a contagious disease to see if they become sick. There is no self-quarantine. Quarantine is a stricter and controlled isolation. Every quarantine is isolation, but every isolation is not quarantine.

Some of the answers to these questions were published online in the two databases *Jezični savjetnik* ('language advice') and *Bolje je hrvatski* ('better in Croatian'[17]), and a special issue of the journal *Hrvatski jezik* (*Croatian language*) on Coronavirus and e-learning during the lockdown. In the papers, the linguistic aspects of COVID-19 pandemic were analyzed from different points of view: cognitive linguistics, i.e., Coronavirus-related metaphors (cf. Štrkalj Despot 2020: 1–7), onomastics[18] – names of people fighting COVID-19 (cf. Vidović 2020: 18–19), phraseology (cf. Kovačević 2020: 25–29), standardology (cf. Blagus Bartolec 2020: 30–32),

16 The speakers asked similar questions about the terms *antigen* ('antigen') and *antitijelo* ('antibody'), *pandemija* ('pandemic') and *epidemija* ('epidemic), *virulencija* ('virulence') and *patogenost* ('pathogenicity'), *cijepiti* and *procijepiti* (both derivatives of *vaccinate* meaning 'to vaccinate and to obtain a vaccination rate'), etc.

17 A website that suggests Croatian equivalents for Anglicisms.

18 Onomastic analysis of the names of people leading the fight against COVID-19 in Croatia.

etymology (cf. Ivšić Majić 2020: 43–44), e-learning (cf. Hudeček/Mihaljević 2020b: 13–17 and Jozić *et al*. 2020: 20–24).

After the publication of the *Glossary of Coronavirus* in April 2020, questions and suggestions from the *Glossary* users followed. They suggested the inclusion of some terms, proposed new words (*drive-in* > *dovozni*), and asked for etymological, normative, and pragmatic explanations (why is it correct to spell *koronavirus* and not *korona virus*) and suggested corrections or changes of some terms and/or definitions. They offered praise (well done for *drive in* > *provozni*) and criticism e.g., *zaštitna maska* ('protective mask') is not the same as *kirurška maska* ('surgical mask'), *imunološki sustav* should be *imunosni sustav* ('immune system').

Some of the comments are shown in Table 12.

Table 12: Questions on the *Glossary of Coronavirus*.

Question	Answer
Is it correct to say *epidemiološka* or *epidemijska kriza* 'epidemiological crisis'?	The word *epidemijski* is derived from *epidemija* ('epidemic') and the word *epidemiološki* is derived from *epidemiologija* ('epidemiology'). So the adjective *epidemijski* is related to the *epidemic*, and the adjective *epidemiološki* is related to *epidemiology*. An epidemic causes the crisis, so the term should be *epidemijska kriza* ('epidemic crisis').
Shouldn't *epidemija* ('epidemic') be replaced by the Croatian word *pošast*?	The word *pošast* is a synonym of *epidemija*, so *epidemija* can be replaced by *pošast* in the general language. However, *epidemija* is a medical term and *pošast* belongs only to the general language.
Could the Croatian neologism *raskuživalo* replace *dezinficijens* ('disinfectant')?	The Croatian term *raskuživač* has the same meaning as *raskuživalo* so there is no need to form a new term from the same stem with a different suffix.
Is it better to say *simptomski bris* or *simptomatski bris* ('symptomatic swab')?	As the adjective should relate to the noun *simptom* ('symptom') according to Croatian word-formation rules, it should be *simptomski bris*.

6 Coronavirus-related neologisms in Croatian lexicography

Many countries have online glossaries and dictionaries of Coronavirus-related terminology e.g., German *Neuer Wortschatz rund um die Coronapandemie*; Dutch *Coronawoordenboek*.

Coronavirus-related vocabulary presented challenges to lexicographers as they had to make many choices in a short time and without (or before) an adequate corpus. “It is a rare experience for lexicographers to observe an exponential rise in usage of a single word in a very short period, and for that word to come overwhelmingly to dominate global discourse, even to the exclusion of most other topics. *Covid-19*, a shortening of *coronavirus disease 2019*, and its various manifestations has done just that.” (cf. OED Blog 2020).

6.1 *Glossary of Coronavirus* – a descriptive dictionary

From the beginning of the pandemic in Croatia (February 2020), lexicographers from the Institute of Croatian Language and Linguistics started collecting Coronavirus-related words and expressions used in Croatian media, social media, and government briefings. They regularly followed daily newspapers *Jutarnji list* and *Večernji list*, TV news, press releases of the Civil Protection Headquarters, and some portals. As there was no Croatian corpus that included Coronavirus-related terms, the terms were collected manually by seven collaborators. The first version of the *Glossary of Coronavirus*[19] was published in the daily newspaper *Jutarnji list* on March 16, 2020. In April 2020, the supplemented version of the *Glossary of Coronavirus* was posted online. In the *Glossary*, for each term a short definition is given. The phrase: *x has the same meaning as y* connects synonymous terms e.g., terms *imunitet krda* (‘herd immunity’) and *kolektivni imunitet* (‘collective immunity’). The *Glossary* included some names (and even nick names) e.g., *Stožer civilne zaštite* (‘civil protection headquarters’), *HZJZ*, acronym of *Hrvatski zavod za javno zdravstvo* (‘Croatian institute for public health’). The purpose of the *Glossary* was to meet the needs of Croatian speakers as soon as possible. It usually records terms as they are used and does not give any normative advice. It includes jargon words as well as scientific terms which entered the general language. Table 13 shows selected entries from the Croatian *Glossary of Coronavirus*. As the *Glossary of Coronavirus* is monolingual we added the English translation.

In October 2021, the *Glossary* had 168 terms. For this paper, we continued collecting new terms in a Google document (so not all terms mentioned in this paper have been recorded in the *Glossary*) and hope that this *Glossary* will be supplemented in the future. After the information on the *Glossary* was published on Facebook of the Institute of Croatian Language and Linguistics[20] (which has 7,700

---

19 The initiative for compiling the *Glossary* came from the principle of the Institute of Croatian Language and Linguistics Željko Jozić. Terms were selected by Goranka Blagus Bartolec, Lana Hudeček, Kristian Lewis, Ivana Matas Ivanković, Maja Matijević, and Milica Mihaljević and the editors of the *Glossary* were Lana Hudeček, Željko Jozić, Kristian Lewis, and Milica Mihaljević.

20 https://hr-hr.facebook.com/ihjj.hr.

Table 13: Selected entries from the *Glossary of Coronavirus*.

Croatian term	Definition	Synonym	English term	English definition
karantena	prisilno, zakonski ili službeno izdvajanje potencijalno zaraženih osoba iz okoline ili čitavih naselja kako bi se spriječilo daljnje širenje virusa; potječe od talijanske riječi *quarantina* ('40 dana'). Prvi put kao zaštitna mjera primijenjena je odlukom Velikoga vijeća Dubrovačke Republike 1377. godine.		quarantine	forced, legal, or official removal of potentially infected persons from the environment or entire settlements to prevent further spreading of the virus; it comes from the Italian word *quarantina* (40 days). It was first applied as a protective measure by the Grand Council of the Republic of Dubrovnik in 1377.
kirurška maska	navlaka koja se stavlja na lice kao zaštita od prenošenja virusa (kašljanjem, izdisanjem, kihanjem i sl.)	zaštitna maska	protective mask	cover that is put on the face as protection against virus transmission (coughing, exhaling, sneezing, etc.)
kliconoša	osoba ili životinja koja u sebi nosi određeni mikroorganizam koji prouzročuje bolest, ali nema znakova bolesti		carrier	a person or animal that carries a particular microorganism that causes the disease but has no signs of the disease
kolapsologija	u novije vrijeme popularan pokret koji upozorava na moguću propast (kolaps) društava te proučava uzroke i posljedice urušavanja civilizacija		collapsology	in recent times, a popular movement that warns of the possible collapse of society and studies the cause and consequences of the collapse of civilization
kolektivni imunitet	imunitet koji je steklo dovoljno ljudi da se bolest prestane širiti	imunitet krda	collective immunity	immunity acquired by enough people to stop the disease from spreading
komorbiditet	istodobna pojava dviju ili više bolesti		comorbidity	simultaneous occurrence of two or more diseases
korona	kolokvijalni naziv za koronavirusnu bolest, tj. COVID-19		Corona	colloquial name for Coronavirus disease, i.e., COVID-19
koronabolnica	bolnica predviđena za smještaj i liječenje zaraženih koronavirusom		COVID hospital	a hospital designed to accommodate and treat people infected by Coronavirus

followers) we received many comments and suggestions from followers. The information on the *Glossary* had 11,619 page views; its reach was 9,921 and impression 736. It had 132 reactions from the users.

6.2 *Croatian Web Dictionary – Mrežnik* – a normative dictionary

After the compilation of the descriptive *Glossary of Coronavirus*, it was decided that some of the Coronavirus-related terms should be included in the normative *Croatian Web Dictionary – Mrežnik*, an online, free, corpus-based, monolingual, hypertext, searchable, normative dictionary compiled at the Institute of Croatian Language and Linguistics.[21] *Mrežnik* has three modules: for schoolchildren, for adult native speakers of Croatian, and for non-native speakers (cf. Hudeček/Mihaljević 2020a). The microstructure of *Mrežnik* (module for adult native speakers) is shown in Figure 1.

The inclusion of Coronavirus-related terms in *Mrežnik* was motivated by the questions and comments from the *Glossary* users, which reflected their need for further explanation and normative guidance.

The selection of terms was made by the editors according to these criteria:
1. more frequent words;
2. words which belong to the standard language;
3. words with a normative or pragmatic problem connected to their usage;
4. Coronavirus-related medical terms used in the general language;
5. word-formation clusters;
6. the most common jargon words;
7. names are not included.

Coronavirus-related terms with compounds and derivatives were added to the *Mrežnik* wordlist, as shown in Table 14. In many cases, these are medical terms that existed in Croatian terminology but would not have been included in a general dictionary if their frequency in everyday discourse did not increase due to COVID-19.

The problem for the Croatian dictionary compilers was that most Coronavirus-related terms could not be found in general Croatian corpora *Croatian Web Corpus* hrWaC and *Croatian Web Repository*.

This meant that the lexicographers had to select examples manually. The website of the *Croatian Web Archive* (Hrvatski arhiv weba) has a thematic collection *COVID-19* in which the lexicographers could check each term and find examples.

However, in October 2021 a small Coronavirus corpus (called *Korona*) consisting of a little over 280,700 words was compiled by Štrkalj Despot and Ostroški Anić. The *Mrežnik* authors and editors used *Korona* corpus in editing Coronavirus-related

21 A demo version of *Mrežnik* (from A to F) is available online on https://rjecnik.hr/mreznik/.

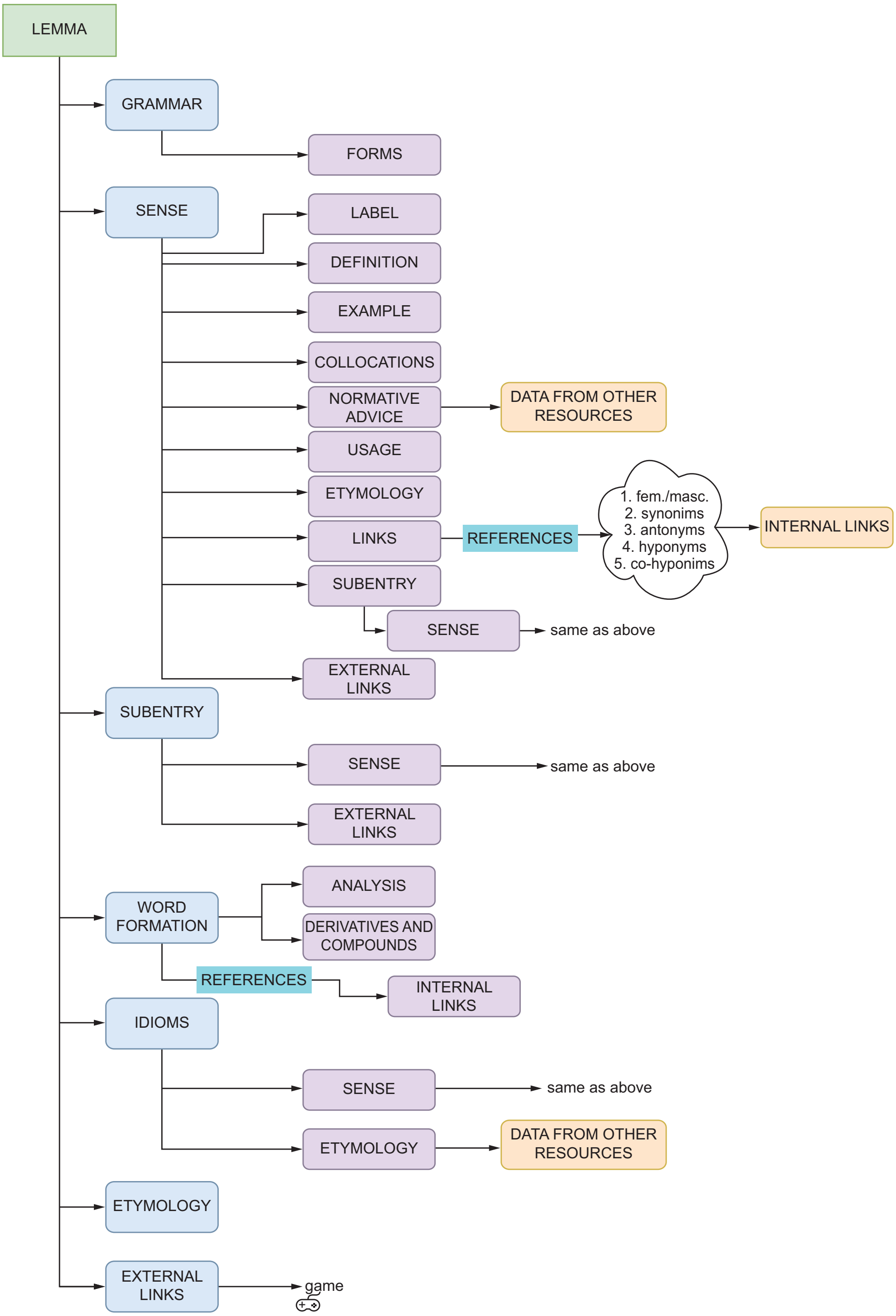


Figure 1: Microstructure of *Mrežnik* – module for adult native speakers.

entries, for supplementing the wordlist (using keywords and N-grams), adding collocations (using word sketches) and examples (using concordance).

Table 14: A sample of Coronavirus-related terms with compounds and derivatives added to the wordlist.[22]

Croatian	English	Croatian	English	Croatian	English
epidemija	epidemic (noun)	dezinfekcija	disinfection	korona	Corona
epidemiolog	epidemiologist	dezinficijens	disinfectant	koronabolnica	Corona hospital
epidemiologinja	epidemiologist (female)	dezinficirati	disinfect	koronafobija	Corona phobia
epidemiologinjin	belonging to the epidemiologist (female)			koronahumor	Corona humor
epidemiologov	belonging to the epidemiologist			koronavirus	Coronavirus
epidemiološki	epidemic (adjective)				

Table 15 shows the entry *COVID-19*[23] with its subentries *COVID ambulanta* ('COVID infirmary'), *COVID bolnica* ('COVID hospital'), *COVID infekcija* ('COVID infection'), *COVID odjel* ('COVID ward'), *COVID ordinacija* ('COVID practice'), *COVID pacijent* ('COVID patient'), and *COVID pozitivan* ('COVID positive'). The terms are illustrated by examples and collocations and the etymology and usage of the term *COVID-19* is explained.

Table 15: Part of the entry *COVID-19* in *Mrežnik*.

Section	Croatian	English
definition	COVID-19 bolest je koju prouzročuje novi soj koronavirusa SARS-CoV-2 (teški akutni respiratorni sindrom, koronavirus 2).	COVID-19 is a disease caused by a new strain of the Coronavirus SARS-CoV-2 (severe acute respiratory syndrome, coronavirus 2).

22 On compiling the wordlist for *Mrežnik* cf. Hudeček/Mihaljević 2020a.
23 The entry *COVID-19* is available on https://rjecnik.hr/mreznik/index.php/covid-19/.

Table 15 (continued)

Section	Croatian	English
corpus example[24]	Bolest COVID-19 prenosi se kapljičnim putem i izravnim kontaktom, preko kapljica sline ili sluzi govorom, disanjem ili kašljanjem zaražene osobe u blizini druge zdrave osobe.	COVID-19 disease is transmitted by droplets and direct contact, through droplets of saliva or mucus by speech, breathing, or coughing of an infected person near another healthy person.
collocations	Kakav je COVID-19? *dugotrajan, težak* Što COVID-19 može? *prijetiti, širiti se, ubiti, uništiti (gospodarstvo, turizam), vladati* Što se s COVID-om 19 može? *biti pozitivan na njega, oboljeti od njega; preboljeti ga* U vezi s COVID-om 19 spominje se: *bolest, borba, kriza, simptom, slučaj*	What is COVID-19 like? *longlasting, difficult* What can COVID-19 do? *threaten, spread, kill, destroy (economy, tourism), rule* What can be done with COVID 19? *be positive about it, get sick of it; get over it* COVID 19 is mentioned in connection with: *illness, struggle, crisis, symptom, case*
subentries	COVID ambulanta COVID bolnica COVID infekcija COVID odjel COVID ordinacija COVID pacijent COVID pozitivan	COVID infirmary COVID hospital COVID infection COVID department COVID practice COVID patient COVID positive

Entries in *Mrežnik* contain collocations introduced by collocational questions and introductory phrases (cf. Hudeček/Mihaljević 2020c: 78–111). At the beginning of the compilation process, as word sketches could not be used for Coronavirus-related collocations, the compilers had to find collocations searching the Internet and *Croatian Web Archive*. However, after the appearance of the new *Korona* corpus, some new collocations were added to the Coronavirus-related entries. Each subentry also has examples and collocations.

Internal links connect synonyms, antonyms, hyponyms, and feminine/masculine pairs.[25] Some of the definitions of Coronavirus-related terms in *Mrežnik* are taken from the *Glossary*. The terms that are recorded in the *Glossary* are linked with external links.

24 In *Mrežnik*, 4–5 corpus examples are given for each meaning of the headword. In Table 13, only one example is given for illustration.

25 Linking is important in *Mrežnik*. Internal links link one *Mrežnik* entry to another (synonyms, antonyms, male/female pairs, word-formation, etc.). External links link a *Mrežnik* entry to an external source (cf. Hudeček/Mihaljević 2019), e.g. the *Glossary of Coronavirus*.

The definitions and linking of two partly synonymous entries are shown in Table 16.

Table 16: Definition and links in the entries *SARS-CoV-2* and *Coronavirus*.

	SARS-CoV-2		Coronavirus	
entry section	**Croatian**	**English**	**Croatian**	**English**
definition 1	SARS-CoV-2 (kratica od engl. *Severe acute respiratory syndrome coronavirus 2* ('teški akutni respiratorni sindrom koronavirus 2') jedan je od sojeva virusa iz istoimene porodice koji prouzročuje koronavirusnu bolest, tj. COVID-19.	SARS-CoV-2 (short for English *Severe acute respiratory syndrome coronavirus* 2) is one of the virus strains from the family of the same name that causes Coronavirus disease, i.e. COVID-19.	Koronavirus jedan je od sojeva virusa iz istoimene porodice koji prouzročuje koronavirusnu bolest, tj. COVID-19.	Coronavirus is one of the virus variants from the family of the same name that causes Coronavirus disease, i.e. COVID-19.
references synonym	koronavirus 1		SARS-CoV-2	
definition 2	–		samo u množini: Koronavirusi su porodica virusa (*Coronaviridae*) koji imaju jednolančani DNK i lipidnu ovojnicu u obliku krune; mogu zaraziti ptice i sisavce, a u posljednje vrijeme i ljude.	only in the plural: Coronaviruses are a family of viruses (*Coronaviridae*) with single-stranded DNA and a crown-shaped lipid envelope; they can infect birds, mammals, and humans.

The headword *SARS-CoV-2* is marked with the terminological label *med*. (medical term). It has a definition and etymological explanation, followed by examples and a link to the synonym *koronavirus* ('Coronavirus'), also a headword in *Mrežnik*. The headword *koronavirus* ('Coronavirus') has two meanings, and only the first is linked to *SARS-CoV-2*.

Normative and/or pragmatic advice is given in all cases where the user might not be sure which word to use or when to use a specific term. Examples of normative and pragmatic advice in *Mrežnik* are shown in Table 17.

Sometimes, new meanings, examples, and collocations have been added to the existing entries (which were not Coronavirus-related) e.g., in the entry *val* ('wave') collocations *drugi val* ('second wave'), *treći val* ('third wave'), and *četvrti val* ('fourth wave') were added. In the entry *balon* ('balloon'), the meaning 'protective measure in which a group of people physically interact or socialize only with each other' and some examples from the Internet and/or corpus were added.

Anosmija/anozmija ('anosmia') is one of the medical terms the frequency of which has increased due to Coronavirus. This term was recorded in two forms *anosmija* and *anozmija*. In the Croatian terminological database *Struna* (cf. Bratanić/Ostroški Anić 2013), the term *anozmija* was preferred.

However, after the appearance of Coronavirus and the entrance of this term into general discourse it was used in the form *anosmija* by the media. Therefore it was decided to have the term *anosmija* as the headword in *Mrežnik* and *anozmija* as its synonym.[26] In the normative note, the reasons for this are explained. The entry *anosmija* in *Mrežnik* is shown in Table 18.

26 The entry *anosmija* is available on https://rjecnik.hr/mreznik/index.php/anosmija/.

Table 17: Normative and pragmatic advice.

Headword	Croatian	English
koronavirus (Coronavirus)	**Normative advice** U stručnoj i široj općoj uporabi u hrvatskome su jeziku ustaljeni prilagođeni nazivi za viruse. Prema pravilu o pisanju stručnih naziva malim se početnim slovom pišu svi jednorječni nazivi te svi višerječni nazivi osim riječi koja je i sama ime ili posvojni pridjev izveden od osobnoga imena, npr. *ebola, zika / virus zika, virus Zapadnoga Nila / zapadnonilski virus*. U skladu s navedenim pravilom piše se i naziv virusa čiji je latinski naziv *Coronavirus* koji je uzročnik aktualne epidemije.	In scientific and general usage, the names of viruses are adapted to the Croatian language. According to the orthographic rule about writing scientific terms, all one word and multi-word terms that do not contain a name or possessive adjective derived from a personal name are written with small initial letters e.g., *ebola* ('ebola')*, zika* ('zika')*, virus zika* ('zika virus')*, virus Zapadnoga Nila / zapadnonilski virus* ('the West Nile Virus'). Following this rule, the term denoting the virus called in Latin *Coronavirus*, which caused the pandemic in 2020 and 2021, is spelled *koronavirus*.
	Pragmatic advice Često se čuje da je netko obolio od koronavirusa. U medijskim se izvještajima katkad govori o broju zaraženih, katkad o broju oboljelih. Oboje je pravilno, ali nije pravilno reći da je netko obolio od koronavirusa. Ljudi nisu oboljeli od virusa nego su njime zaraženi, a od bolesti obolijevaju. Možemo reći da je netko zaražen virusom ili da je obolio od koronavirusne bolesti ili od bolesti COVID-19.	One often says that someone has been sick with Coronavirus. In the media, we can sometimes read about the number of infected or the number of diseased. Both expressions are correct, but it is not right to say that someone is diseased with Coronavirus. Somebody cannot be diseased with a virus only infected by it, and diseased with a disease. We can say that a virus has infected somebody or that somebody is sick with the Coronavirus disease or COVID-19.
epidemija (epidemic)	Epidemija označuje naglo širenje neke bolesti. Stoga, nije pravilno govoriti o pandemiji koronavirusa, nego treba govoriti o epidemiji koronavirusne bolesti, epidemiji bolesti COVID-19 ili u razgovornome stilu o epidemiji korone.	The epidemic denotes a widespread occurrence of an infectious disease. Thus it is not correct to say Coronavirus epidemic. It is right to say an epidemic of Coronavirus disease or an epidemic of COVID-19 or in the colloquial style epidemic of Corona.
pandemija (pandemic)	Pandemija označuje širenje neke bolesti na više država, kontinenata ili na cijeli svijet. Budući da pandemija označuje 'širenje bolesti', nije pravilno govoriti o pandemiji koronavirusa, nego treba govoriti o pandemiji koronavirusne bolesti, pandemiji bolesti COVID-19 ili u razgovornome stilu o pandemiji korone.	The pandemic denotes the spreading of disease on more countries, continents, or the whole world. Thus it is not correct to say Coronavirus pandemic. It is right to say pandemic of Coronavirus disease or pandemic of COVID-19 or in the colloquial style pandemic of Corona.

Table 18: Entry *anosmija* in *Mrežnik*.

Entry section	Croatian	English
definition	Anosmija je gubitak njuha, jedan od rjeđih simptoma bolesti COVID-19, najčešće se pojavljuje u mladih pacijenata s blagom kliničkom slikom.	Anosmia is the inability to smell, one of the less common symptoms of COVID-19 disease, most often occurring in young patients with a mild clinical picture.
synonym	anozmija	
pragmatic note	U medicinskome nazivlju upotrebljava se i naziv *anozmija* te je u bazi *Strune anozmija* preporučeni medicinski naziv, a *anosmija* dopušteni. U medijima se (povezano s pandemijom bolesti COVID-19) upotrebljava naziv *anosmija*.	The term *anosmia* is used in medical terminology, and the medical term is recommended in the *Struna* database, and *anosmia* is an allowed term. The term *anosmia* is used in the media (associated with the COVID-19 pandemic).
external links	to *Struna* database to *Glossary of Coronavirus*	

7 Conclusions

Coronavirus has influenced many walks of life, from health, medicine, sociology, and psychology to education and language. Languages change over time, but in 2020 and 2021, change had been more rapid than ever. The appearance of Coronavirus resulted in the formation of many new words and phrases, the evolution of meaning of the existing ones, and the passing of terms from the language of science to everyday discourse. The media and social media had a crucial role in the creation and spreading of these new terms.

As in many other languages, the beginning of the COVID-19 pandemic in Croatia was associated with the appearance of many new words (*koronabolnica* ‘Corona hospital’, *superširitelj* ‘super-spreader’) and phrases (*socijalna distanca* ‘social distance’) as well as additional meanings of the existing ones (*balon* ‘balloon’, *val* ‘wave’). Native speakers, especially journalists, often asked linguists from the Institute of Croatian Language and Linguistics for language advice. This was the reason for language advice posted on the portals *Jezični savjetnik* and *Bolje je hrvatski!*, for a special issue of the journal *Hrvatski jezik* connected with Coronavirus and e-learning, for the compilation of the descriptive *Glossary of Coroanvirus*, and the inclusion of Coronavirus-related neologisms in the *Croatian Web Dictionary – Mrežnik*. In autumn 2021, Coronavirus-related terms were not yet included in a general publicly available corpus of the Croatian language, but a small specialized *Korona* corpus had been

compiled and was available on request. Corona-related neologisms were also not yet included in the then most recent Croatian dictionary of neologisms *Rječnik neologizama u hrvatskome* (cf. Muhvić-Dimanovski/Skelin Horvat/Hriberski 2016).

Some of the new Coronavirus-related words/phrases/meanings still belong only to the spoken jargon, and they will probably never enter the Croatian standard language (*kovidiot* 'covidiot'). On the other hand, some Coronavirus-related terms already belong to the standard language (*koronabolnica* 'Corona hospital'). This new terminology presented numerous challenges for linguists and especially for standardologists and lexicographers. Standardologists were faced with many questions relating to orthography, grammar (morphology, syntax, and especially word-formation) as well as to the lexis posed by journalists and general language users.

Bibliography

Alyeksyeyeva, I. O./Chaiuk, T. A./Galitska, E. A. (2020): Coronaspeak as Key to Coronaculture: Studying New Cultural Practices through Neologisms. In: *International Journal of English Linguistics* 10/6, 202–212.

Blagus Bartolec, G. (2020): Jezik u doba korone. In: *Hrvatski jezik* 7/2, 30–32.

Bratanić, M./Ostroški Anić, A. (2013): The Croatian National Termbank STRUNA: A New Platform for Terminological Work. In: *Collegium antropologicum* 3/37, 677–683.

Brozović Rončević, D./Ćavar, D. (2008): Hrvatska jezična riznica kao podloga jezičnim i jezičnopovijesnim istraživanjima hrvatskoga jezika. In: Petrović, B./Samardžija, M. (eds.): *Vidjeti Ohrid: referati hrvatskih sudionica i sudionika za XIV. međunarodni slavistički kongres*. Zagreb: Hrvatsko filološko društvo/Hrvatska svečilišna naklada, 173–186.

Hudeček, L./Mihaljević, M. (2019): *Croatian Web Dictionary – Mrežnik* – Linking with Other Language Resources. In: Kosem, I. *et al.* (eds.): *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference*. Brno: Lexical Computing CZ, 72–98.

Hudeček, L./Mihaljević, M. (2020a): The *Croatian Web Dictionary – Mrežnik* Project – Goals and Achievements. In: *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 46/2, 645–667.

Hudeček, L./Mihaljević, M. (2020b): Videolekcije za maturante – kako izbjeći najčešće jezične pogreške u školskome eseju. In: *Hrvatski jezik* 7/2, 13–17.

Hudeček, L./Mihaljević, M. (2020c): Collocations in the *Croatian Web Dictionary – Mrežnik*. In: *Slovenščina 2.0*. 8/2, 78–111.

Ivšić Majić, D. (2020): Otrovna kruna je u narodu. In: *Hrvatski jezik* 7/2, 43–44.

Janusheva, V. (2020): The Macedonian Language in Regard to Covid-19. In: *Thesis* 9/2, 243–261.

Jozić, Ž. *et al*. (2020): Mrežni izvori u vrijeme korone. In: *Hrvatski jezik* 7/2, 20–24.

Karachina, O. E. (2020): Language of Pandemic: Linguistic and Cultural Aspects. In: *Russian Linguistic Bulletin* 2/22, 45–48.

Khalfan, M./Batool, H./Shehzad, W. (2020): Covid-19 Neologisms and Their Social Use: An Analysis from the Perspective of Linguistic Relativism. In: *Linguistic and Literature Review* 6/2, 117–129.

Kilgarriff, A. et al. (2004): The Sketch Engine. In: Williams, Geoffrey/Vessier, Sandra (eds.): *Proceedings of the 11th EURALEX International Congress*. Lorient: Universite de Bretagne-Sud, 105–116.

Kovačević, B. (2020): Između čekića i plesa. In: *Hrvatski jezik* 7/2, 25–29.
Lewis, K. (2021): Od krave do EMA-e: cijepljenje i cjepivo u jezičnome okružju. In: *Hrvatski jezik* 8/2, 28–33.
Ljubešić, N./Klubička, F. (2016): Croatian web corpus hrWaC 2.1. Slovenian language resource repository CLARIN.SI. [http://hdl.handle.net/11356/1064; last access: October 26, 2021].
Miloš, I. (2020): Pandemija viceva u pandemiji korone: zaraza koju volimo. In: *Hrvatski jezik* 7/3, 9–11.
Muhvić-Dimanovski, V./Skelin Horvat, A./Hriberski, D. (2016): *Rječnik neologizama u hrvatskome jeziku* [http://rjecnik.neologizam.ffzg.unizg.hr/; last access: October 26, 2021]
Štrkalj Despot, K. (2020): Kako koronavirus mijenja jezik kojim govorimo (i mislimo)?. In: *Hrvatski jezik* 7/2, 1–7.
Štrkalj Despot, K./Ostroški Anić, A. (2021): A War on War Metaphor: Metaphorical Framings in Croatian Discourse on Covid-19. In: *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 47/1, 1–36.
Vidović, D. (2020): Hrabro, ali postupno, mudro i oprezno – o osobnim imenima i prezimenima predvodnika borbe s koronavirusom. In: *Hrvatski jezik* 7/2, 18–19.

Internet sources

anosmija. *Hrvatski mrežni rječnik* – Mrežnik. *Osnovni modul*. Institut za hrvatski jezik i jezikoslovlje. [https://rjecnik.hr/mreznik/index.php/anosmija/ last access: October 29, 2021].
Andrews, A. (2020): *The Coronavirus Recession is a "She-cession"*. [https://msmagazine.com/2020/05/15/covid19-women-unemployment-economy-coronavirus-recession-is-a-she-cession/; last access: October 26, 2021].
Bolje je hrvatski. [http://bolje.hr, last access: June 24, 2021].
Coronaspeak – A Guide to the Language We Are All Using These Days. [https://grownandflown.com/corona-slang-language-guide/, last access: June 24, 2021].
Coronavirus: Glossary of Common Terms. [https://www.webmd.com/lung/coronavirus-glossary#1, last access: June 24, 2021].
Coronawoordenboek. [https://www.taalbank.nl/2020/03/14/coronawoordenboek/; last access: June 24, 2021].
COVID-19. *Hrvatski mrežni rječnik* – Mrežnik. *Osnovni modul*. Institut za hrvatski jezik i jezikoslovlje. [https://rjecnik.hr/mreznik/index.php/covid-19/, last access: October 26, 2021].
Croatian Web Archive (Hrvatski arhiv weba). [https://haw.nsk.hr/tematske-zbirke/24/covid19-1; last access: June 24, 2021].
English Web 2020 (enTenTen20). [https://www.sketchengine.eu/ententen-english-corpus/; last access: October 24, 2021].
Hrvatski mrežni rječnik – Mrežnik (from A to F). [https://rjecnik.hr/mreznik/; last access: October 24, 2021].
Hrvatsko društvo za primijenjenu lingvistiku, HDPL konferencija 2021, Program skupa. [http://www.hdpl.hdpl.hr, last access: October 24, 2021].
Jezični savjetnik. [http://jezicni-savjetnik.hr/; last access: June 24, 2021].
Lawson, R. (2020): *Coronavirus has led to an explosion of new words and phrases – and that helps us cope*. [https://theconversation.com/amp/coronavirus-has-led-to-an-explosion-of-new-

words-and-phrases-and-that-helps-us-cope-136909%26%23x003F last access: October 26, 2021].
Macmillan Dictionary. [https://www.macmillandictionary.com/us/dictionary/american/covidiot last access: October 24, 2021].
Mahdawi, A. (2020): *From Covidiot to doomscrolling how coronavirus is changing our language*. [https://amp.theguardian.com/commentisfree/2020/apr/15/from-covidiot-to-doomscrolling-how-coronavirus-is-changing-our-language?/; last access: October 26, 2021].
Mali rječnik Instituta za jezik – koronakriza, infodemija, koronašoping, samoizolacija, koronamanija . . . Kako nam je pandemija Covida-19 u samo mjesec-dva promijenila svakodnevni jezik. Jutarnji list. [https://www.jutarnji.hr/vijesti/hrvatska/koronakriza-infodemija-koronasoping-samoizolacija-koronamanija-kako-nam-je-pandemija-covida-19-u-samo-mjesec-dva-promijenila-svakodnevni-jezik-10113136; last access: June 24, 2021].
Neuer Wortschatz rund um die Coronapandemie. [https://www.owid.de/docs/neo/listen/corona.jsp#; last access: June 24, 2021].
Neuer Wortschatz rund um die Coronapandemie. [https://www.owid.de/docs/neo/listen/corona.jsp#; last access: June 24, 2021].
OED Blog. 2020. Social change and linguistic change: the language of Covid-19. [https://public.oed.com/blog/the-language-of-covid-19/; last accessed: October 26, 2021].
Pojmovnik koronavirusa. [https://jezik.hr/koronavirus; last access: June 24, 2021].
SCIMETH. *Suvremene metode obrade jezika*. [https://cji.uniri.hr/scimeth-2021/, last access: June 24, 2021].
Struna. [http://struna.ihjj.hr/; last access: July 26, 2021].
Urban Dictionary. [https://www.urbandictionary.com/define.php?term=Quarantime; last access: October 26, 2021].
Words We're Watching: 'Coronial'. Merriam-Webster. [https://www.merriam-webster.com/words-at-play/meaning-of-coronial-covid-kids; last access: October 26, 2021].

Sílvia Barbosa, Susana Duarte Martins

The neologisms of the COVID-19 pandemic in European Portuguese: From media to dictionary

1 Introduction

Humanity has already been confronted with pandemics in the past, such as the Spanish Flu (1918), the Asian Flu (1957), AIDS (1981), H1N1 (2009), Ebola (2014), and Zika Virus (2015), just to name a few. COVID-19 (2020), however, has had a stronger impact on the lives of people around the world, so that communication about new treatments, new care, new concerns, new behaviours was necessary and, whenever a discovery of this social and clinical reality happened new words and/or neological expressions emerged at an extremely fast pace, "simultaneously a manifestation of language evolution and the evolution of knowledge" (Lino 2019: 10).

While adjusting to the COVID-19 pandemic, people around the world started to talk about the "new normal" way of life, and they conveyed feelings and thoughts on the topic through social networks and traditional communication channels resorting to a set of specific linguistic strategies, such as metaphors and neologisms.

The vocabulary in different domains and in everyday speech was expanded to accommodate a complex social, cultural, and professional phenomenon of changes. Therefore, this new life gave birth to a new language – the "coronaspeak".

According to Thorne (2020), the "coronaspeak" has three stages: first, it emerged in the way medical aspects were communicated in everyday language; secondly, it occurred when speakers verbalized the experiences they had undergone and "invented their own terms"; finally, this "new" way of speaking emerged in the government and authorities' jargon, to ensure that the new rules and policies were understood, and that population adopted socially responsible behaviours.

In this paper, we will focus on the second stage, because we intend to take stock of how speakers communicate and verbalize this new way of living, particularly on social networks, for example. Alongside, we are interested in the context in which the neologism – be it a new word, a new meaning, or a new use – emerged, is used,

Acknowledgements: This paper is supported by the Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia as part of the project Centro de Linguística da Universidade NOVA de Lisboa – UID/LIN/03213/2020.

Sílvia Barbosa, NOVA CLUNL, NOVA University of Lisbon, e-mail: silviabarbosa@fcsh.unl.pt
Susana Duarte Martins, NOVA CLUNL, NOVA University of Lisbon, e-mail: susanaduartemartins@fcsh.unl.pt

https://doi.org/10.1515/9783110798081-010

and understood, through the observation of the occurrence of the new word(s) either on social networks or through dissemination texts (press) to confront it with the ones that Portuguese digital dictionaries have attested so far. Different criteria regarding the insertion of new units, the inclusion date, and the lexicographic description of the entries in the dictionaries will be debated.

2 Neology: theoretical and methodological core issues

Historically, neologisms have been the target of prejudice and stigmatization when confronted with the standard language (Boulanger 2010). Although speakers usually recognize the units of their language that may be considered as new, the concept of "neologism" is debatable. Back in 1976, Rey wondered whether "neologism" is a concept or simply a pseudo-concept, arguing that "there is obviously no neologism per se, but in relation to a set of arbitrarily defined uses" (1976: 17). For this linguist, the concept of neologism is methodological and pragmatic.

Given the complexity of the concepts of neology (the process) and neologisms (the product), experts have divergent opinions about this topic. Therefore, this phenomenon has been examined from different angles over time, such as the studies carried out by Guilbert (1975), Rey (1976), Alves (1990), Cabré, Freixa and Solé (2002), Pruvost and Sablayrolles (2003), or Boulanger (2010). More recently, we highlight the works of Alves and Maroneze (2018), Jesus (2018), and Rio-Torto's (2020) analysis of the lexical renovation in Brazilian and European Portuguese.

According to Alves and Maroneze (2018: 9), the challenge in defining neologism resides in the concept of novelty: "New word regarding what or whom?" When we speak about novelty, a new concept arises in the scope of lexical innovation: the "novelty feeling", a criterion of psychological nature, that Guilbert (1975: 31) associates with neologisms to express the way speakers may experience the designation of new concepts. On the other hand, Guerrero Ramos (2017) and Lino (2019) defend that the "neological feeling" is crucial to identify and delimit a neologism, despite its fluctuating character (Sablayrolles 2006).

Pruvost and Sablayrolles (2003) consider neologisation as a natural process, which depends on several factors, such as age, the speakers' experience, and the dynamic of different periods, as it is the case of the global pandemic of COVID-19. Following Correia and Lemos (2005), we understand neology as (i) the natural ability to renew the lexicon of a language by creating and incorporating new units; and (ii) the study (observation, collection, description, and analysis) of neologisms that occur in linguistic systems.

Neologisms start by appearing in the speech, and may eventually become fixed in the language, thus losing their neologism status. As Guilbert puts it: "the repetition of

the act of creation establishes the individual neologism in the lexicon society; creation is confirmed by a certain usage. The created term is then lexicalised and, at the same time, loses its neologism quality[1] to become a socially established word" (1975: 49). Hence, all the words were once neologisms, as Jean-Claude Boulanger and Bernard Quemada defend (Sablayrolles 2016).

Some authors also mention the distinction between occasionalisms and neologisms (Dal and Namer 2016, Bueno Ruiz 2020) based on the stability function of a unit in a linguistic system and its permanence time (Dressler 1981). While an occasionalism would be temporary and ephemeral, without dictionary attestation, a neologism, from a diachronic point of view, is a unit that can be included into a dictionary at a given moment, and thus become part of the common lexicon of a language losing its neologic status.

The entry into the linguistic system, made official by the registration in the language dictionary, of permanent and stable formations, resulting from a system need, mainly of denominative character, coincides with the moment when those units cease to be neologisms, according to Correia (1998). In this sense, Guerrero Ramos (2017: 1399) claims that a unit is neological if it does not appear in the dictionary, therefore "the dictionary remains an effective means of measuring neology". On the other hand, media plays an important role in the identification of neologisms as a vehicle for the dissemination of the standard language and is considered a reference model that can condition or encourage the use of certain linguistic trends (Freitas, Ramilo, Arim 2010) or, as Pruvost and Sablayrolles (2003: 9) state: "Press chronicles, more or less selective institutions and dictionaries also play their regulatory role to evaluate, channel, define, suggest, sometimes officially impose an adaptation or substitutes for the neologisms resulting from the daily turmoil".

Given the sophistication of neologisms, experts have come forward with many proposals to categorize neological units. Sablayrolles (1997) addresses this issue, presenting 12 types of neologisms typologies alongside formation processes. Despite the numerous existing typologies, Jesus (2018: 54) concludes that "in general, it is possible to synthesize the formation processes in three fundamental types: formal, semantic and loan processes",[2] and advocates the importance of considering "pragmatic and discursive factors as inherent to any neological unit and, consequently, to any typologisation proposal".

1 In the original text: "qualité de néologisme".

2 Formal neologisms are creations based on processes of derivation, composition, formation by acronyms, reduction of words or even in the creation of innovative roots (Boulanger 1979). In this group, Rey (1976) includes some borrowings, along with ex nihilo creations, morphological units, initialisms, and acronyms. When a new meaning is given to a form that already exists in the language, the neologism is semantic and can be described by different types of novelty: total or partial (Rey 1976). When the neological units derive from the adoption of a foreign unit, they are known as borrowed neologisms.

3 Neology in Portugal in a nutshell

In Portugal, neologism studies began in the 1980s, at NOVA CLUNL ('Linguistics Research Centre of NOVA University of Lisbon'), with the *Observatoire du Français Contemporain de Lisbonne* ('Observatory of Contemporary French in Lisbon'), under the supervision of Teresa Lino. Neologisms were collected and processed as computerized data in the *Base de Neologismos do Português Contemporâneo* ('Contemporary Portuguese Neologisms Base') (Lino 1988, 2003). Later, the *Observatório do Português Contemporâneo* ('Contemporary Portuguese Observatory') was founded with the aim of creating a bank of Portuguese neologisms, with the same methodological principles as the French observatory, followed then by the *Observatório de Neologia e de Terminologia em Língua Portuguesa* – Neoporterm ('Observatory of Neology and Terminology in Portuguese Language'). In 2004, the *Observatório de Neologia* – ONP ('Neology Observatory') was created by Margarita Correia at ILTEC (currently CELGA-ILTEC, 'Center for the Study of General and Applied Linguistics'), as a part of the observatories' network of the *Neologia das Línguas Românicas* – NEoRom ('Neology of Romance Languages'), a project coordinated by Teresa Cabré (Correia *et al.* 2006).

Despite several advances in the neology work in Portugal, both projects (at NOVA CLUNL and CELGA-ILTEC) are on standby.

4 European Portuguese dictionaries

The different evolution stages traversed by dictionaries over time contemplate "an attitude towards language(s) and a reflection on language(s) itself (themselves): the dictionary has been a cultural object ever since it was created and has strived to define the lexical corpus of a language with a descriptive, didactic, and sometimes normalizing perspective", as Lino (2018: 609) clarifies. The dictionary has an extremely important social role in the preparation of an individual for society. It is an example of balance between the correct use of language and its variation, whether dialectal or orthographic and a guarantor of lexical reliability, without which no norm or rule survives.

In Portugal, there is no normative language policy, nor an institution with legal competence to determine the linguistic norm, despite the existence of the Academia das Ciências de Lisboa ('Lisbon Academy of Sciences'), which provides the Portuguese government with consultancy in linguistic and scientific matters of national interest. Consequently, there are no normative dictionaries, unlike what happens, for example, in Spain, where one can find several scientific academies (Real Academia Española, 'Royal Spanish Academy'; Institut de Estudis Catalanes, 'Institute of Catalan Studies'; Real Academia Gallega, 'Galician Royal Academy'; Euskaltzaindia,

'Royal Academy of the Basque Language') with legal powers to determine the norm of languages, and normative dictionaries are available to the public.

The European Portuguese dictionaries only have a descriptive status and, in some cases, a status of reference dictionaries, which allows the regulation of linguistic products despite not having legal competence to establish the standard. However, the speakers acknowledge Portuguese dictionaries as reliable sources of the intended correct uses of the language (Correia 2009).

Usually, paper-based dictionaries have an introduction, a preface and/or a user's guide, where questions regarding objectives, methodology, macro and microstructure, the insertion of new lexical units, number of entries, among others may be addressed. The kind of lexicographic information shared with users, as well as the extension of it, depends on the dictionary's scope. Regarding the new entries, and since a unit ceases to be neological once it is attested in a dictionary, Correia (1998) strongly disagrees with the inclusion of the label "neol.", often used in the microstructure of paper-based dictionaries to indicate the most recent units, defending that it is "theoretically more correct, to choose the date of the first attestation of the registered unit" (idem: 62). The *Dicionário do Português Atual Houaiss* – DPAH (2011, 'Houaiss Dictionary of the Portuguese Current Language') includes both the dates of the entries and the label "neol.".

The Portuguese online dictionaries: *Dicionário Infopédia da Língua Portuguesa* (DILP, 'Portuguese Language Infopedia Dictionary') and *Dicionário Priberam da Língua Portuguesa* (DPLP, 'Priberam Dictionary of the Portuguese Language') do not explicitly mention the objectives of the works, the methodology used, or details about how the insertion of new lexical units is done, nor whether there is a temporal register associated with new entries. However, the DPLP provides more lexicographic information (e.g., number of entries) and has a detailed user guide[3] when compared with the DILP.

Unlike what happens with paper-based dictionaries, where the different editions serve as a timestamp to identify diachronic changes, it is difficult to verify aspects such as the insertion date of a neologism, or the introduction of new meaning through reformulation or adaptation of definitions in the Portuguese e-dictionaries, given that the lexicographic criteria governing the insertion of units are unclear.

Therefore, in order to find out when a new unit is included in the Portuguese digital dictionaries, we have to perform a diachronic analysis of their entries. Taking COVID-19 as an example, we have evidence from both dictionaries of the introduction of at least one new related word between December 2020 and April 2021.

In 2020, there were two units associated with the new pandemic in the DILP (Figure 1).

3 Available on: https://dicionario.priberam.org/consultar.aspx.

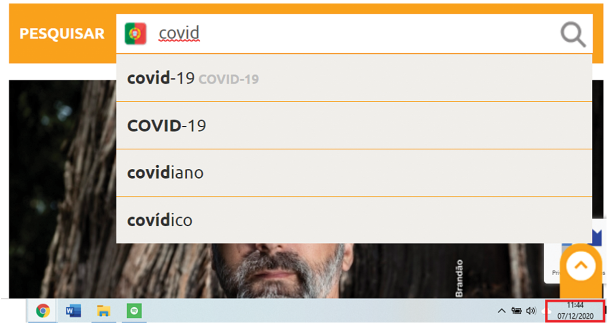


Figure 1: Entries associated with COVID-19 in the DILP on 07/12/2020. Source: DILP.

In 2021, *covidário* comes up as a new unit in the DILP (Figure 2).

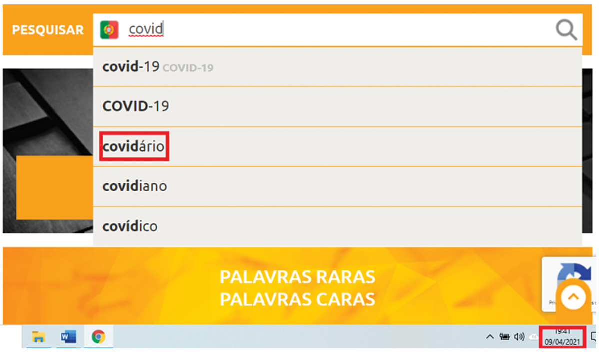


Figure 2: Entries associated with COVID-19 in the DILP on 09/04/2021. Source: DILP.

In 2020, *covidário* was already an entry of the DPLP, as we can observe in Figure 3.



Figure 3: Entries associated with COVID-19 in the DPLP on 07/12/2020. Source: DPLP.

In 2021, the DPLP has added *covid-drive* to its lemma list (Figure 4).

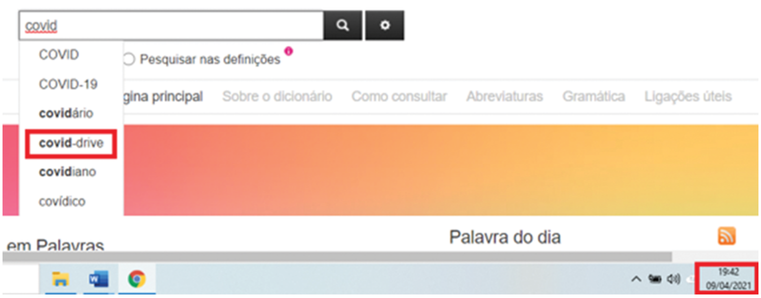


Figure 4: Entries associated with COVID-19 in the DPLP on 09/04/2021. Source: DPLP.

From the observation of the two dictionaries, at first glance, we could say that the DPLP seems to have different criteria from the DILP, allowing a faster insertion of new units: the entry *covidário* was already in the DPLP when the DILP included it in its lemma list.

Despite these constraints, the DILP and DPLP are two digital reference works for the contemporary Portuguese language and, therefore, a reliable source of comparison between the units attested in the dictionaries and the neological units registered in media and social networks, since dictionaries also use them as part of their corpora.

4.1 *Dicionário Infopédia da Língua Portuguesa* – DILP

The DILP has both a paper-based and a digital version, the latter one is incorporated in the *Infopédia* website: www.infopedia.pt/.

In the Infopédia page, we have access to many linguistic resources: monolingual dictionaries,[4] bilingual dictionaries,[5] a multimedia encyclopedia, a Portuguese orthographic vocabulary, a spell converter, translation and spelling games, trivia: “rare words, expensive words”, language doubts, and the “word of the year”, allowing users to send comments or suggestions about any resource.

---

4 Eleven Portuguese language dictionaries: Portuguese language dictionaries with and without spelling agreement, Portuguese sign language dictionary, dictionary of Portuguese verbs, dictionary of acronyms and abbreviations, toponymy dictionary, dictionary of proper names (anthroponymy), dictionary of medical terms, dictionary of Latin phrases and foreign expressions, basic illustrated dictionary, dictionary of Portuguese for foreigners.

5 Nine bilingual dictionaries: Chinese, Dutch, English, French, German, Greek, Italian, Spanish, Tetum, and two verb dictionaries of English and French.

The DILP does not provide any official information about the total number of entries, however, the corresponding paper version – *Dicionário Moderno da Língua Portuguesa* (2008, 'Modern Dictionary of the Portuguese Language'), belonging to the publishing house Porto Editora, records "32,500 entries, phrases and idioms, providing more than 2,700 examples and 66,500 definitions"[6] and, in an announcement regarding the launch of the digital version of the dictionary in 2007, the publishing house informed that the DILP had more than 240,000 definitions available,[7] so we know that at the time the DILP's lemma list had fewer entries than its paper-based counterpart. The Porto Editora commercializes a higher number of paper-based dictionaries (including more thematic/technical dictionaries) and other products (grammar books and handbooks) than the ones available on the Infopédia website.

The microstructure of the DILP comprises phonetic transcription, syllabic division, etymology, grammatical information, usage marks (e.g. colloquial, regionalism, taboo, slang, other linguistic varieties of Portuguese, etc.), synonyms, antonyms, related words and anagrams, foreign words, idioms, some examples or expressions to illustrate contexts. The entries are also attached to information in sign language and references to articles from other dictionaries or its encyclopedia. In the *Dicionário de Português para Estrangeiros* – DPE (2020, 'Dictionary of Portuguese for Foreigners'), users can also listen to the words' pronunciation.

Despite the omission of the total number of the DILP lemmas, as regular users, we have the perception of an increase of available lemmas, justified by the creation of new dictionaries in recent years, like the DPE. The Infopédia enables access to lots of linguistic information through numerous resources. Besides that, it is user-friendly and has an inviting image.

4.2 *Dicionário Priberam da Língua Portuguesa* – DPLP

The DPLP only has an online version, available through Priberam website: https://dicionario.priberam.org/, along with other resources, like FLiP[8] and LegiX.[9]

In the Priberam webpage, we have access to several linguistic resources: translation assistants (English, French, Spanish); verb conjugator (European/Brazilian Portuguese and Spanish); spelling agreement converter; syntactic and spell checker

---

6 <https://www.portoeditora.pt/produtos/ficha/dicionario-moderno-da-lingua-portuguesa/200124>; last access: August 5, 2021.

7 <https://www.portoeditora.pt/noticias/dicionario-da-lingua-portuguesa-gratuito-na-internet/759>; last access: August 8, 2021.

8 Several linguistic resources for the Portuguese language: https://www.flip.pt/, last access: August 8, 2021.

9 Portuguese legal databases, chosen by the largest law firms operating in Portugal.

(European/Brazilian Portuguese and Spanish); grammar; vocabulary, with two distinct lexical bases for European Portuguese and Brazilian Portuguese.

The DPLP is a lexicographic product that resulted from the paper dictionary *Novo Dicionário Lello da Língua Portuguesa* (1996–1999, 'New Lello Dictionary of the Portuguese Language'). It contains 133,000 entries, "including phrases and phraseologies, whose lemma list comprises the general vocabulary and the most common terms of the main scientific and technical areas",[10] according to the introduction of the dictionary on its webpage.

The dictionary allows users to customize it according to the desired Portuguese variety: European or Brazilian standards, and one can also choose to use the spelling agreement ('acordo ortográfico') version or not,[11] depending on specific needs.

After setting the preferences, the DPLP offers users options of autocomplete search and spell check, and the entries include information about the morphological analysis, search (cross-reference) in the definitions, verb conjugator, related words, translation assistants, similar and nearby words (or neighbouring words), the occurrence of the unit in other entries, as well as the real use of the word in blogs, media, and Twitter. It also contains data concerning etymology and pronunciation cues, grammar information, usage marks (e.g., colloquial, informal, regionalism, slang, other linguistic varieties of Portuguese, etc.), synonyms and antonyms, anagrams, foreign words, idioms, contexts, between others.

Having a quick look at the pros and cons of Priberam, all the information displayed in the DPLP is free, except for lexicographic resources associated with FLiP. It is a user-friendly resource, and there is a tutorial that guides us along the different sections and types of searches that the dictionary enables. If we are seeking informal uses of language and contexts of words, the DPLP is a good choice, since it shows us the word in examples taken from blogs, media, and Twitter, and allows the user to have more up-to-date information regarding contexts of informal uses of language, specifically slang, when compared to the DILP.

---

10 <https://dicionario.priberam.org/sobre.aspx>; last access: August 5, 2021.

11 The "new" spelling agreement (also known as *spelling agreement of 1990*) is mandatory in Portugal, Brazil, Cape Verde, São Tomé and Príncipe, but still under discussion in Angola and Mozambique. As for Guinea-Bissau, Equatorial Guinea and East-Timor, the governments' priority is that the population speak the official language (Portuguese), and with education/linguistic policies to implement the Portuguese language in the country comes the spelling agreement (if communities have access to student books, grammars, and other linguistic resources from countries where the spelling agreement is applicable, or through teachers from those countries).

5 Methodology

In this work, we intend to: (i) verify which lexical units emerged relating to the COVID-19 period in the media and social networks, and from these units which ones were included into a dictionary at a given moment; and (ii) observe which lexical units were attested by the dictionaries in a very short period to meet the users' needs.

The neologisms candidates were extracted from media (newspapers *Público* and *Expresso*) and social networks (such as Facebook and Twitter) in a period comprised between December 2019 – considered as the beginning of the pandemic, and July 2021.

Considering Sablayrolles statement that "a word enters a dictionary because it is no longer neological" and "a word is neological because it is not in the dictionary" (2006: 141), we have selected the two digital dictionaries of European Portuguese mentioned above – DILP and DPLP, to confirm if the extracted units are neologisms or not.

These dictionaries were chosen for the following reasons: they (i) are freely available (despite being owned by two publishing houses: Porto Editora and Priberam), (ii) possess a considerable lemma list, (iii) are representative of the lexicon of European Portuguese, (iv) are targeted to the Portuguese and lusophone audience.

5.1 Criteria for the selection of neologisms candidates

The process of identifying and collecting neologisms candidates was carried out in two stages, manual and semi-manual, and was based on four criteria: diachronic, psychological, systematic instability, and lexicographic (Cabré 1993), which helped us to delimit the units considered as neologisms candidates.

Following Cabré (1993), a unit is neological if it has appeared recently (diachronic criterion). On the other hand, as Guilbert (1975) declares, a unit is felt as new at a given moment by the speakers of a particular linguistic community compared to the language stage immediately before (psychological criterion), whether it is a new orthographic entity, a new meaning, or an update of meaning. This criterion, dubbed as "novelty characteristic", is responsible for the immediate delimitation of a candidate.

Cabré (1993) refers to the formal instability of the neologism as relevant to its classification, a unit will be considered neological if, cumulatively, it shows signs of morphological, phonetic, or spelling instability. Different spellings and graphic markings, hesitation concerning grammatical gender or pronunciation, for example, reveals insecurity towards the use or existence of specific units in the language.

Finally, we follow the lexicographic criterion, in which the candidate unit will be neological if it is not yet registered in a language dictionary, either at the level of the entry or meaning. The inclusion of the unit into a dictionary at a given moment reveals that it has lost its neological nature.

5.2 Selection of neologisms candidates

Once the criteria were defined, we have started collecting candidates. As mentioned above, since the Portuguese neology observatories are on standby, the solution was to endeavour a manual or semi-manual selection. Although time-consuming, it was considered indispensable, and only so could we successfully identify cases of semantic and formal neology (Correia & Lemos 2005).

First, whenever a unit "felt" to be neological was found, it was introduced in the candidate list with the information considered relevant in the candidate form, as shown in Table 1:

Table 1: Candidates related to *álcool* and *gel*.

candidate	N.o	context	source	date
álcool-gel (*alcohol-gel*)	0038	A distância das mesas, a separação dos irmãos nos recreios, as setas no chão, os frascos de **álcool-gel** e até o ecrã da televisão com os números da covid-19 aparecem desenhados.	Newspaper Expresso	2020/09/29
álcool gel (*alcohol gel*)	0032	Metro e autocarros cheios de regras e de lugarers vazios, viajantes mascarados cobertos de respeito, fiscais de poucas multas e de muita pedagogia, ou máquinas automáticas com chocolates, pastilhas elásticas, máscaras, luvas e **álcool gel**. Eis o retrato dos transportes públicos do Porto, onde um autocarro seguia sem passageiros para Sonhos.	Newspaper Expresso	2020/03/19
álcool em gel (*alcohol in gel*)	0019	A LMVH proprietária da Louis Vuitton, começou a produzir e a distribuir gratuitamente **álcool em gel**, em França. A iniciativa procura responder ao risco de falta de desinfectantes, em todo o país, para proteger a população da propagação do novo coronavírus.	Newspaper Público	2020/03/19
gel desinfe(c)tante (*disinfectant gel*)	0006	Os receios face ao vírus quadriplicaram as vendas de Fevereiro de sabão e **gel desinfectante**, incluindo aqueles com base de álcool, face ao ano passado, segundo os dados mostrados pela gigante do retalho Lotte Mart.	Newspaper Público	2020/03/04

From the final extraction, we have obtained a list of candidates, of different typologies and formation processes: *álcool-gel* ('alcohol-gel'), *ano pandémico* ('pandemic year'), *antigo normal* ('old normal'), *antimáscaras* ('antimasks'), *bolhas domésticas* ('domestic bubbles'), *centro de vacinação* ('vaccination center'), *comportamentos de risco* ('risk behaviours'), *confinamento* ('lockdown'), *desconfinamento* ('lifting

lockdown'), *drone pandémico* ('pandemic drone'), *escola virtual* ('virtual school'), *estado de emergência* ('emergency state'), *fase de mitigação* ('mitigation phase'), *fraudemia* ('fraudemic'), *geração pandemia* ('pandemic generation'), *hidroxicloroquina* ('hydroxychloroquine'), *imunidade de grupo* ('group immunity'), *janelas do confinamento* ('lockdown windows'), *kit de diagnóstico* ('diagnostic kit'), *língua covid* ('covid language'), *mapa pandémico* ('pandemic map'), *negacionista* ('negationist'), *plano nacional de testagem* ('national testing plan'), *quarentena* ('quarantine'), *recém-vacinados* ('newly vaccinated'), *supercontagiadores* ('super contagious'), *testes serológicos* ('serological tests'), *uberização* ('uberization'), *vacinódromo* ('vaccinedrome'), *zaragatoa* ('swab').

Subsequently, candidates were lemmatized to facilitate their registration and allow a more efficient analysis. For reasons of space and scope, we will only discuss the results derived from the selected units: *coronavírus*, *COVID-19*, *pandemia* and the prefix *tele-*. These units were chosen due to the simple identification of the candidate, and their associations with the disease, such as different units related to the designation of the disease (*coronavírus*, *COVID-19*), units used as a metonym for a specific disease (*pandemia*), and a prefix unit (*tele-*) associated with performing certain tasks in the so-called "new normal" or "post-pandemic scenario".

6 Analysis

In this section, we are going to dwell on the lexicographic representation of four units related to the pandemic. The selection was based on the most generic units analyzed when the subject is the COVID-19 pandemic (*coronavirus, COVID-19, pandemic*). Additionally, the particular interest in understanding the impact of COVID-19 on people's lives in the technological age justifies our choice of the prefix *tele-* as a potential promoter of neologisms.

Initially, we confirmed whether the four selected units were included in the lemma list of the DILP and DPLP. Subsequently, we observed the microstructure of the entries, namely the content of the definitions and the type of words related to the units under study in both dictionaries (5.1.). In the final stage of our research, we present the neologisms candidates collected from different sources and discuss their formation processes (5.2.).

6.1 Lexicographic description of the units: *coronavirus, COVID-19, pandemic,* and the prefix *tele-*

We will set our attention to the four selected examples, compare their definitions in the DILP and DPLP and discuss the approach to the same units in both dictionaries.

After commenting on the definition of the units under study, a comparative analysis of the microstructure of both dictionaries was carried out as a result of two types of search processes. First, we have performed a search in the lookup window of each dictionary, where a set of words starting with the same characters (e.g. "coronav") is displayed (Figure 5).

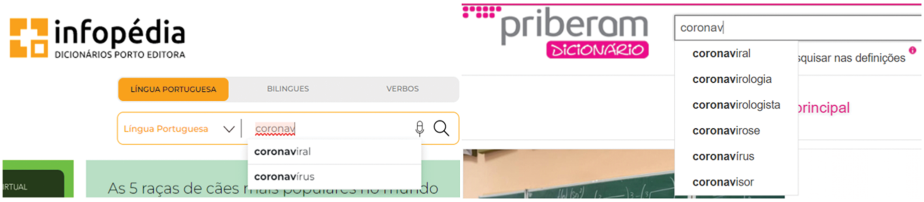


Figure 5: Example of a search result with the initial characters of coronavirus in the lemma list of the DILP (left) and DPLP (right).

Then, we have observed the related words within each dictionary entry under study (Figure 6).



Figure 6: Example of the entry and respective suggestions of other words in the DILP (left) and DPLP (right).

The lookup window of the DILP displays other words in alphabetical order (up to 10 results), and the page of the entry allows users to see a set of related words ('veja também', "see also"). In the DPLP, the lookup window suggests other words in alphabetical order (up to 6 results), while on the page of the entry, we have related words ('palavras relacionadas'), similar words ('parecidas'), nearby/neighbouring words ('palavras vizinhas').

In a preliminary analysis, we have checked the suggestions in the lookup window and compared them with the related words for each of the four entries in both dictionaries between April and July 2021. During the analysis, we have discarded words that are not semantically related or, for the sake of length, words that show up

simultaneously in the two types of searches or if they occur with other units under analysis simply because the lemma occurs within the definition of another entry (as in the DPLP section "this word in the dictionary"). Ultimately, this diachronic analysis has not retrieved significant differences in the referred period, so we will only present a summary (*cf*. Tables 2–5) for each unit with a summary of our findings.

6.1.1 coronavírus

The lemma *coronavírus* is included in both dictionaries, and their definitions include data about etymology, gender (it is a masculine noun, with no graphic variation nor gender instability), number (singular and plural), and it is associated with the domains of medicine and biology (Figure 7).

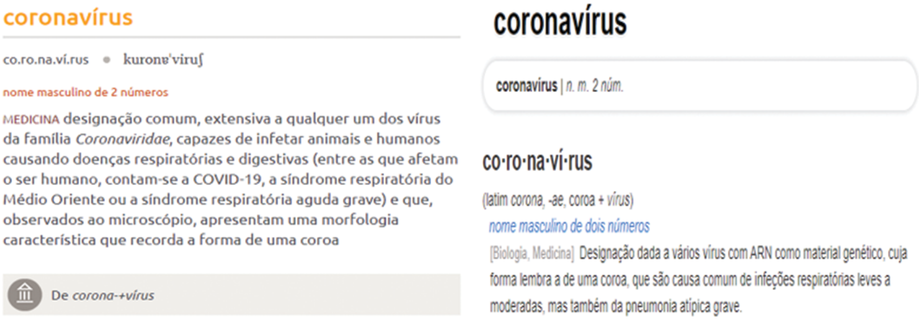


Figure 7: The entry of the lemma coronavírus in the DILP[12] (left) and DPLP[13] (right).

The definition of *coronavírus* in both dictionaries stresses the following characteristics of the lemma: (i) it is a common designation of a certain family of viruses; (ii) the virus causes a set of symptoms; (iii) it has the shape of a crown ("coroa", "corona" in Latin).

Additionally, the DILP includes encyclopedic information in the definitory text, naming different types of *coronavírus*: COVID-19, MERS-CoV (acronym of Middle East respiratory syndrome coronavirus, or in Portuguese: "síndrome respiratória do Médio

12 DILP definition: 'common designation, extended to any of the viruses of the *Coronaviridae* family, capable of infecting animals and humans, causing respiratory and digestive diseases (among those that affect humans, there are COVID-19, the Middle East respiratory syndrome or the severe acute respiratory syndrome) and which, when viewed under a microscope, have a characteristic morphology reminiscent of the shape of a crown.'.

13 DPLP definition: 'Designation given to several viruses with RNA as a genetic material, whose shape resembles a crown, which are a common cause of mild to moderate respiratory infections, but also of severe atypical pneumonia.'.

Oriente") and SARS (acronym of Severe Acute Respiratory Syndrome, in Portuguese: "síndrome respiratória aguda grave"). The fact that this definition includes the recent coronavirus disease (COVID-19), suggests a reformulation or an update of the definition to accommodate a new concept.

Let us have a look at the data retrieved from both dictionaries (Table 2) regarding the related and nearby/neighbouring words (*cf*. Figure 7):

Table 2: Data retrieved from the DILP and DPLP regarding *coronavirus* related and nearby words.

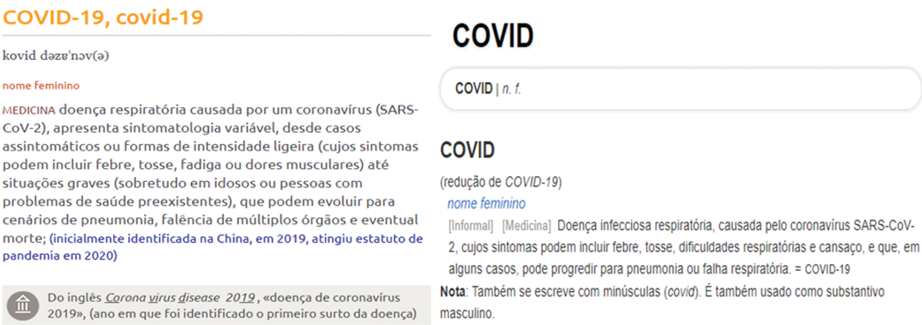
14 units retrieved from coronavírus	DILP	DPLP
alfacoronavírus	√	√
betacoronavírus	√	√
coroa	√	√
corona	√	√
coronafobia	√	**X**
coronafóbico	√	**X**
coronaviral	√	√
coronavirologia	**X**	√
coronavirologista	**X**	√
coronavirose	**X**	√
coronavírus	√	√
coronavisor	**X**	√
deltacoronavírus	√	√
gamacoronavírus	√	√
Total of results	10	12

From the comparison of the units retrieved from *coronavírus* in both dictionaries, one can remark that the DPLP seems to have more *coronavírus* related entries than the DILP. On the other hand, the DPLP includes units that have no definition available – *coronavirose* ('canine and feline coronavirus') and *coronavisor* (possibly referring to a face shield for corona, even though we are using often "viseira" to designate the same device), so if we were to consider words with definitions, we can say that the DPLP and DILP are even.

Finally, we observe that *coronavirose*, *coronavisor*, *coronavirologia* ('coronavirology') and *coronavirologista* ('coronavirologist'), as well as *coronafobia* ('coronaphobia') and *coronafóbico* ('coronaphobic'), show traces of instability concerning their fixation as entries in both dictionaries. Therefore, we believe that these units are losing their neologism status as they are experiencing a process of being included into a dictionary at a given moment.

6.1.2 COVID-19

The DILP displays in a single entry *COVID-19* and *covid-19*. On the other hand, the DPLP attests two entries for the same concept: *COVID* and *COVID-19*, however, the latter has no definition attached and users are informed that *COVID-19* is not in the dictionary and are invited to suggest the "inclusion of the searched word in the dictionary". Until very recently, the lemma *COVID-19* had a definition in the DPLP, now it appears solely in the entry of *COVID* as a reduction and synonym of *COVID-19*, alongside the observation that it can also be spelt in small caps: *covid*.


COVID-19, covid-19

kovid dəzɐ'nɔv(ə)

nome feminino

MEDICINA doença respiratória causada por um coronavírus (SARS-CoV-2), apresenta sintomatologia variável, desde casos assintomáticos ou formas de intensidade ligeira (cujos sintomas podem incluir febre, tosse, fadiga ou dores musculares) até situações graves (sobretudo em idosos ou pessoas com problemas de saúde preexistentes), que podem evoluir para cenários de pneumonia, falência de múltiplos órgãos e eventual morte; (inicialmente identificada na China, em 2019, atingiu estatuto de pandemia em 2020)

Do inglês *Corona virus disease 2019*, «doença de coronavírus 2019», (ano em que foi identificado o primeiro surto da doença)

COVID

COVID | *n. f.*

COVID

(redução de *COVID-19*)

nome feminino

[Informal] [Medicina] Doença infecciosa respiratória, causada pelo coronavírus SARS-CoV-2, cujos sintomas podem incluir febre, tosse, dificuldades respiratórias e cansaço, e que, em alguns casos, pode progredir para pneumonia ou falha respiratória. = COVID-19

Nota: Também se escreve com minúsculas (*covid*). É também usado como substantivo masculino.

Figure 8: The entry of the lemma COVID-19 in the DILP[14] (left) and DPLP[15] (right).

Looking at both definitions, we notice that the two dictionaries favour the female gender, despite a note in the DPLP regarding the possibility of the occurrence of the masculine gender, a tendency also observed in other Latin languages. Besides gender instability, this lemma presents spelling variants concerning the use of uppercase or lowercase, uppercase being the preferred form in both dictionaries. The lemma is categorized as a noun, and it derives from an English acronym, a piece of information disregarded in the DPLP since the lemma etymology is absent. It is associated with the domain of medicine, yet the entry *COVID* in the DPLP adds that this shorter form of *COVID-19* is also informal.

These definitions emphasize the following characteristics about the lemma: (i) the year of the outbreak (only in the DILP, as there was an update on the lemma

14 DIPL definition: 'respiratory disease caused by a coronavirus (SARS-CoV-2), which presents variable symptoms, from asymptomatic cases or forms of mild intensity (whose symptoms may include fever, cough, fatigue or muscle pain) to severe situations (especially in the elderly or people with pre-existing health problems), which can evolve into scenarios of pneumonia, multiple organ failure and eventual death; (initially identified in China in 2019, it has reached the pandemic status in 2020).'.

15 DPLP definition: 'An infectious respiratory disease caused by the SARS-CoV-2 coronavirus whose symptoms may include fever, cough, breathing difficulties, and tiredness, and which in some cases may progress to pneumonia or respiratory failure.'.

definition in the DPLP); (ii) the type of illness: a respiratory (infectious, in the definitory text of the DPLP) disease, with symptoms (mild to severe); (iii) the seriousness of the illness: respiratory failure (in the DPLP), eventual death (in the DILP); (iv) the cause of the disease: SARS-CoV-2 coronavirus; (v) the origin of the disease (China) and its status (pandemic status in 2020), only in the DILP.

Concerning the lexicographic representation of *SARS-CoV-2*, we verify that, while the *SARS* lemma is included in a dictionary at a given moment, the specific version of coronavirus – *SARS-CoV-2* is not, despite being included in the definitory text of both dictionaries. The fact that *SARS-CoV-2* is a medical term, with an acronymic basis, more complex in terms of structure and, even, verbalization in European Portuguese, may explain the preference for the form *COVID-19* or, simply, *COVID* in current language dictionaries. When we look up *SARS-CoV-2* in the *Dicionário de Termos Médicos* ('Dictionary of Medical Terms'), the search panel cross-references us immediately to *COVID-19*, the definition in which the term is used. On the other hand, when we look up *SARS-CoV* in the same dictionary, no results are returned, but if we check the entry of *SARS*, we realize that *SARS-CoV* is part of its definition. Thus, there is a lack of standardization in the search results that are returned by the dictionary to its users.

As for data regarding the related and nearby/neighbouring words (*cf.* Figure 8), we have identified eight units in both dictionaries (Table 3):

Table 3: Data retrieved from the DILP and DPLP regarding *COVID-19* related and nearby words.

8 units retrieved from *COVID-19*	**DILP**	**DPLP**
anticovid	X	√
anticovid-19	√	X
COVID	X	√
COVID-19	√	X
covidário	√	√
covidiano	√	√
covídico	√	√
covid-drive	X	√
Total of results	5	6

The results show that the lexicographers of the DILP and DPLP took different decisions about the treatment of the equivalent units: the DILP includes the entry + a specific element (number): *COVID-19*, *anticovid-19*, while the DPLP presents the same units without that specific element. Also here, the DPLP exhibits slightly higher results than the DILP.

Despite being treated as synonyms in the DILP and DPLP, *covidiano* (‘covidian’) and *covídico* (‘covidic’)[16] may present different semantic values in dictionaries and social networks. In fact, in social networks, *covidiano* shows up in contexts where the lemma is the result of a phonological game related to the nearby Portuguese unit *quotidiano* (‘daily’). As for *covid-drive*, one can say that the lemma is losing its novelty characteristic since it is already part of the lemma list of one of the two dictionaries.

6.1.3 pandemia

The lemma *pandemia* is part of the DILP’s and DPLP’s lemma list, and the entries display information about etymology, gender (it is a feminine noun, showing no graphic variation or gender instability), and it is associated with the domain of medicine in DILP (Figure 9).

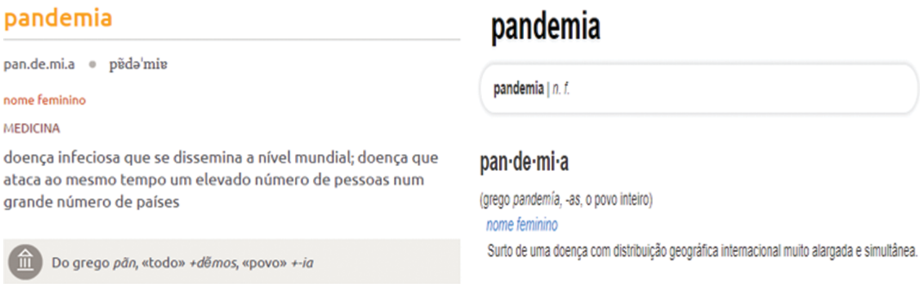


Figure 9: The entry of the lemma pandemia in the DILP[17] (left) and DPLP[18] (right).

The definitions of the *pandemia* lemma highlight the following characteristics: (i) it is an outbreak (DPLP) of an (infectious) disease, (ii) it spreads worldwide, (iii) and affects a high number of people (iv) simultaneously.

The unit *pandemia* entered the Portuguese language in 1873 (DPAH) and occurs 73 times in the CETEMPúblico corpus (data from 1991–1998).[19] All the occurrences are associated with the HIV pandemic, the 1993 Cholera pandemic, and the 1918 Spanish flu pandemic. Comparing our research with these data, one may conclude that *pandemia*

16 Even though Google retrieves 154,000 and 10,300 occurrences of *covidian* and *covidic* (in 15/08/2021), these lemmas are absent from the main online dictionaries of English (*Cambridge*, *Collins*, *Dictionary.com*, *Macmillan*, *Merriam-Webster*, *Oxford*).

17 DILP definition: ‘Infectious disease that spreads worldwide; disease that attacks a large number of people in a large number of countries at the same time.’.

18 DPLP definition ‘Outbreak of a disease with a very wide and simultaneous international geographic distribution.’.

19 <https://www.linguateca.pt/CETEMPublico/>; last access: August 1, 2021.

is used as a synonym of *SARS-CoV-2* (found in 2002), such as *coronavírus*, or *COVID-19*. On the contrary, despite being discovered in 1965, the lemma *coronavírus* is absent of the DPAH and CETEMPúblico.

One interesting remark concerning related and nearby/neighbouring words of *pandemia* is that they differ in the dictionaries under study: in the DILP pandemic is associated with coronavirus, COVID-19, and outbreak (*surto*), while the DPLP connects it to a calamity (*calamidade*), but also fatigue (*fadiga*) or tiredness (*cansaço*) and covidic (*covídico*). Once again, the DPLP presents a few more entries related to *pandemia* than the DILP (Table 4):

Table 4: Data retrieved from the DILP and DPLP regarding *pandemic* related and nearby words.

13 units retrieved from pandemia	DILP	DPLP
calamidade	X	√
cansaço	X	√
coronavírus	√	X
COVID-19	√	X
covídico	X	√
endemia	√	X
epidemia	√	X
fadiga	X	√
pandemia	√	√
pandêmico	√	√
pandemiologia	X	√
pandemiológico	X	√
surto	√	X
Total of results	7	8

The units retrieved from *coronavirus* – *pandemiologia* ('pandemiology') and *pandemiológico* ('pandemiological'), are not found in the DILP, however, the DPLP integrates these units in its lemma list, similarly to what happened with the coronavirus related words with the same suffixes (*coronavirologia* and *coronavirologista*). Additionally, our research showed that even though *pandemiologia* is not included in the DILP, *endemiologia* ('endemiology') is attested in that dictionary. Given the instability of the insertion of these units in the dictionaries, they can also be considered as cases of units losing their neologism status.

6.1.4 tele-

Both dictionaries mention that the prefix *tele-* is a compositional element associated with the concept of *distance*, however, only the DPLP makes explicit that this prefix can also be used as a truncated element of *television* (Figure 10).

Similarly, Cunha and Cintra (1984) attested these two homonymous compositional elements in the context of the prefix *tele-*, both related to distance and television. Since the lemma *teledisco* (music video) is included in both dictionaries and the information regarding its etymology indicates that *tele-* is a truncation of *television* (*tele[visão]+disco*), one can remark that the definition of this prefix is incomplete in the DILP.



Figure 10: The entry of the lemma regarding the prefix -tele in the DILP[20] (left) and DPLP[21] (right).

When it comes to the nearby/neighbouring and related words, the DPLP is much more prolific than the DILP (Table 5):

Table 5: Data retrieved from the DILP and DPLP regarding *tele-* related and nearby words.

14 units retrieved from COVID-19	DILP	DPLP
longe	√	X
teleadministração	X	√
telealarme	X	√
telealuno	X	√
teleautografia	X	√
teleautógrafo	X	√
telecêntrico	X	√
telecomandar	X	√
teledifundir	X	√
teledinâmico	X	√
teledirigir	X	√
teleguiar	X	√
televisual	X	√
telex	X	√
Total of results	1	13

While the DILP only associates the unit *longe* ('far') to the prefix *tele-*, the DPLP relates it to units that convey the concepts of distance (*teleadministração*, 'teleadministration'; *telealarme*, 'telealarm'; *telealuno*, 'telestudent'; *teleautografia*, 'telauthography'; *teleautógrafo*, 'telautograph'; *telecêntrico*, 'telecentric'; *telecomandar/teleguiar*, 'to operate (something) by remote control; to remote-control'; *teledinâmico*, 'teledynamic'; *teledirigir*, 'to control from a long distance'; *telex*) and television (*teledifundir*, 'to broadcast by television'; *televisual*; and *telealuno*, that can also occur in the context of the television) following its definition.

In short, we can conclude that most units related to the chosen examples are included in the lemma list of the DILP and DPLP. As for the few units that only show up in one of the dictionaries, we may infer that undisclosed lexicographic reasons are underlying these decisions, simultaneously demonstrating traces of variability and instability in their being included into a dictionary at a given moment, probably connected with the loss of neologicity.

6.2 Neological creativity

In this section, we will discuss the cases of neologisms collected from various sources, such as media (newspapers, magazines) and social networks (Facebook, Twitter), as well as their typology, given that "the press and the media, in general, are an important gateway not only for common neologisms but also, and even more so, for specialized neologisms" (Guerrero Ramos 2017: 1399).

6.2.1 coronavírus

Graphic variation concerning *coronavírus* and not pointed out in the e-dictionaries was identified in several sources: *Coronavírus*, *corona virus* or the reduction *corona, Corona*. The lemma *corona* is also attested in the DPLP as a synonym of *coronavirus*, used informally. Although *coronavírus* is not understood as a full synonym of *COVID-19*, *novo coronavírus* (new coronavirus) was assumed to be synonymous in some contexts.

Regarding the processes of neological formation, we have identified several cases of neologisms candidates created through means of prefixation, suffixation, compounding, importation of loanwords, as well as syntagmation, as shown below:

(i) prefixation: *pré-corona* ('pre-corona'), *pré-coronavírus* ('pre-coronavirus'), *pós-coronavírus* ('post-coronavirus'), *anti-corona* ('anti-corona');

20 DILP definition "element of word formation that expresses the idea from far, far, at a distance".

21 DPLP definition "1. Expresses the notion of distance (e.g., telecommuting). 2. Expresses the notion of television (e.g., music video)".

(ii) suffixation: *coronado/a* (*corona+-ado/a*, 'with corona', 'coronated'; *corona+lixado/a*, 'angry because of corona'), *coronafobia* ('coronafobia'), *coronafóbico* ('coronafobic');
(iii) compounding: *alfacoronavírus* ('alfacoronavirus'); *betacoronavírus* ('betacoronavirus'); *deltacoronavírus* ('deltacoronavirus'); *gamacoronavírus* ('gamacoronavirus'); *corona bónus* ('corona bonus'); *coronaditadura* ('corona dictatorship'), *corona-histeria* ('corona hysteria'), *coronatroika*, *coronavirose* ('canine coronavirus'), *coronavisor* ('face shields against corona'), *coronaviral* ('coronaviral'), *coronavirologia* ('coronavirology'), *coronavirologista* ('coronavirologist'), *troikavírus* (troika+virus);
(iv) syntagmation (phrasal noun constructions):
   – noun+adjective (corona): *nação corona* ('corona-nation'), *imposto corona* ('corona tax'), *geração corona* ('corona generation'), *presidência corona* ('corona presidency');
   – noun+preposition (*de*)+(article)+noun (corona): *festa do coronavírus* ('coronavirus party'), *festas do corona* ('corona parties'), *tempos de corona* ('corona times');
(v) loanwords: *Corona bonds, corona bonds, coronabonds*, *corona-jihad*, *corona room*, *corona-app*, *corona party*, *corona free*, *coronababies*, *Darth Corona*.

Corona and *coronavirus* are the base of units formed by prefixation, although the base *coronavírus* seems less productive in our corpus. On the other hand, *corona* is categorized both as a noun and an adjective, and it can present two genders: masculine (*o coronavírus*, given the gender of *virus* in Portuguese) or feminine (when adjective of a feminine noun: *a corona geração*). Graphic variation stands out in the loanwords.

6.2.2 COVID-19

COVID-19 is lexicalized as a noun in the European Portuguese dictionaries, as speakers lose awareness of its acronymic origin. In addition to the variation mentioned in the dictionaries (*COVID*-19, *covid-19* and *COVID*, *covid*), other cases of graphic variation were found: *Covid-19*, *CÓVID-19*, *Covid*.

Given the challenges facing the identification of the formation process of some units, in some cases concurrently associated with more than a single process, we have assigned the neologisms to the most obvious word-formation process and have followed the suggestions of the glossary *A covid-19 na língua* ('The Covid-19 in the language', Ciberdúvidas da Língua Portuguesa, 2020).[22] Prefixation, suffixation,

22 <https://ciberduvidas.iscte-iul.pt/artigos/rubricas/idioma/covid-19-na-lingua/4059>; last access: August 16, 2021.

initialisms, blending, syntagmation, and loanwords were the most productive among our analysis:

(i) prefixation: *anticovid-19* ('anti-covid-19'), *eurocovid* ('euro-covid'), *pós-covid* ('post-covid'), *pós-covidologia* ('post-covidology'), *pré-covid* ('pre-covid');
(ii) suffixation: *covidade* (COVID-19 dissemination among sportsmen), *covidário* (place in a health establishment intended for the care and treatment of patients with suspected or confirmed infection by COVID-19), *covidês* (*COVID+-ês*, 'COVID-19 language'), *covidiano*, *covideiro* (a person who does not deny the existence of a pandemic), *covídico* ('covidic');
(iii) compounding: *covidarte* (*COVID+arte*, 'covidart');
(iv) blending: *covidar* (*COVID+convidar* e/ou *conversar*, 'to invite and/or chat'), *covidar-se* (*COVID+infetar-se* ou *convidar-se*, 'become infected or invite oneself'), *covidiário* (*COVID+diário*, 'covidiary'), *covidivórcios* (*COVID+divórcios*, 'covidivorces'), *covidizer*, *covidizerque* (*COVID+que ouvi dizer (que)*, 'that I heard (that)');
(v) initialisms: *a.C.* (*antes da COVID-19*, 'before COVID-19'), d.C. (*depois da COVID-19*, 'after COVID-19');
(vi) syntagmation:
   - noun+adjective (covid): *cães covid* ('covid dogs'), *enfermaria covid* ('covid infirmary'), *fado covid* ('covid fado'), *língua covid* ('covid tongue'), *multa covid* ('covid fine');
   - noun+preposition (*de*)+noun (covid): *pandemia de covid-19* ('COVID-19 pandemic'), *surto de covid-19* ('COVID-19 outbreak'), *vítima de covid-19* ('COVID-19 victim');
   - noun+preposition (*de*)+article+noun (covid): *ditadura da covid-19* ('COVID-19 dictatorship'), *transmissão do covid* ('covid transmission');
   - noun+preposition (*de*)+noun+preposition (*de*)+article+noun (covid): *taxa de transmissão da covid-19* ('covid-19 transmission rate');
   - noun+preposition+article+noun (covid): *vacina contra a covid-19* ('vaccine against the covid-19');
(vii) loanwords: *covidiota* ('covidiot'), *long Covid*, *StayAway Covid*.

The neologisms associated with *COVID-19* also show a high rate of graphic variation. As in *coronavírus*, it can either occur as a noun and an adjective and can present two genders, even though the masculine gender is not considered standard given that the unit is assigned with the gender of *disease*, *a doença* (a feminine gender unit in Portuguese).

6.2.3 Pandemia

Pandemic was the source of many cases of lexical creativity. These neologisms were created mostly through processes of prefixation, suffixation, parasynthesis, blending, and syntagmation, as follows:

(i) prefixation: *antipandémico* ('antipandemic'), *pós-pandemia* ('post-pandemic'), *pré-pandemia* ('pre-pandemic');
(ii) *suffixation: pandémico* (adjective: 'pandemic');
(iii) blending: *fraudemia* (*fraude+pandemia*, 'fraudemic'), *infodemia* (*informação+pandemia*, 'infodemic'), *pãodemia* (*pão+pandemia*, 'breademic');
(iv) parasynthesis: *antipandémico* ('antipandemic'), *pós-pandémico* ('post-pandemic'), *infodémico* (adjective: 'infodemic');
(v) syntagmation:
   - noun+adjective (*pandemia*): *pandemia covid-19* ('covid-19 pandemic'), *geração pandemia* ('generation pandemic');
   - noun+preposition (*de*)+noun: *pandemia de covid-19* ('covid-19 pandemic'), *diário de pandemia* ('pandemic diary'), *tempo(s) de pandemia* ('pandemic time(s)');
   - noun+preposition+article+noun: *combate à pandemia* ('fight the pandemic'), *batalha contra a pandemia* ('battle against the pandemic'), *pico da pandemia* ('pandemic peak'), *propagação da pandemia* ('pandemic spread'), *pandemia da desinformação* ('pandemic misinformation'), *pandemia da pobreza* ('pandemic poverty'), *contenção da pandemia* ('pandemic containment');
   - noun+preposition+article+adjective+noun: *pandemia do novo coronavírus* ('new coronavírus pandemic');
   - verb+(article)+noun: *controlar (a) pandemia*, *gerir a pandemia* ('to control/manage the pandemic');
   - verb+preposition+article+noun: *lutar contra a pandemia* ('fight against the pandemic').

(vi) loanwords: *pand-emmys* (by blending of *pandemic+emmys*).

Contrary to the previous units, graphic variation and gender instability is not characteristic of *pandemic* related neologisms, but similarly, it has been used as a noun and an adjective to form phrasal noun constructions.

6.2.4 Tele-

It seems that the prefix *tele-* was related only with distance during a given time, and later, with the advent of the television, the prefix lexical productivity also became associated with this device. However, with the COVID-19 pandemic, the creation of new units with the semantic value of "distance" took on a prominent role, highlighting society need for social distancing. It is the case of *teletrabalho* ('teleworking',

'telecommuting'), considered by some as a luxury; *burguesia do teletrabalho* ('teleworking bourgeoisie'), regarding the highest-paid people; *telemedicina* ('telemedicine'), specially conceived to remote patients; *telejulgamento* ('telecourt') or virtual trials; *telescola* ('teleschool'), or *telensino* ('teleteaching'), the official designation adopted in Madeira Island. The lexical creativity associated with remote learning is not new. Specifically created for students (*telealunos*, 'telestudents') who lived in isolated locations or were unable to enroll in a school due to lack of vacancies, the *telescola* operated in Portugal between 1965 and 2003. With the arrival of the pandemic, this teaching concept was reactivated and the television programme "Study at home" was created. Consequently, an adjustment in the definition of these units was necessary, given that the *telestudents* can attend lessons not only by means of the television (the only medium available in the past) but also through various devices connected to the Internet.

Neologisms regarding leisure or social activities, like *telepraxe* ('telehazing'), have also emerged in this context of physical distance, and performing tasks remotely through the Internet originated new units, such as *teleconsulta* ('teleappointment'), *telemanutenção* ('telemaintenance'), *teleconsultoria* ('teleconsulting'). The government institutions had to adjust to this new concept of distance, starting to exercise it through *teleadministração* ('teleadministration') to (try to) maintain a *teledemocracia* ('teledemocracy') state. While teleworking, one can still *teledizer mal dos colegas* ('telespeak ill of co-workers') over an online meeting, and if experiencing problems with electronic devices purchased at *Worten*, this Portuguese store can *tele-resolver* ('telesolve') clients' technical issues over the phone.

7 Conclusions

Individuals and institutions, such as national language academies, are responsible for the creation of neologisms, whether in the current language or scientific and technical settings. The COVID-19 outbreak encouraged lexical creativity, facilitating communication regarding individual and social perceptions towards the new life experiences boosted by the pandemic.

Most of the selected units (*coronavírus*, *COVID-19*, *pandemia*, *tele-*) and nearby/neighbouring words are attested in the European Portuguese e-dictionaries (DILP, DPLP). The units that only appear in one of the works demonstrate the variability of the lexicographic criteria of the dictionaries. We have identified entries, considered pertinent enough to integrate the dictionaries lemma list, that are waiting for the inclusion of definitions (DPLP) or entries that cross-referenced users to other entries where the searched unit occurs (DILP). These situations are a novelty in the lexicographic setting, given that they would never happen in paper-based dictionaries. Another interesting aspect for reflection is that even while conducting our research, we

have detected a few changes in the lemma list (*cf.* Figures 1–4) and definitory text of some of the units under study. The attestation of new units in a preliminary phase (in the lemma list or the entry microstructure only including the grammatical category, and the information that the definition will be added soon), may be explained by the lexicographers' will to respond to society's linguistic needs, with sufficient efficiency and speed, typical of digital resources that require constant updates. However, diachronic research would be needed to confirm if the lexicographic description of the entries will be completed (DPLP) or if units occurring in attested entries but not yet included in a dictionary at a given moment will be added (DILP).

As a rule, the Portuguese digital dictionaries do not mention the date of the first occurrence or insertion of the lemmas, nonetheless, the date may occasionally appear within their definitions, as in the case of units concerning particular diseases (*COVID-19*, *cf.* Figure 8). On the other hand, if one needs to monitor the insertion of new units or the reformulation/adaptation of definitions in these dictionaries, virtually the only methodology at our reach is restricted to taking screenshots of the lookup window and of the entries microstructure to observe the lexicographic representation of certain units at a given time. The analysis of the lexicographic representation of the selected units and nearby/neighbouring words led us to the conclusion that no objective or obvious criteria are underlying the insertion of new units in the European Portuguese e-dictionaries, contrary to what happens in the Brazilian Portuguese (digital) dictionaries, like the *Dicionário Caldas Aulete* ('Caldas Aulete Dictionary'), where entries may be labelled as "new", "original" or "updated" entry.

Moving on to the neologisms candidates, it was not an easy task to classify the units according to their formation processes as the literature confirms. On the other hand, pragmatic and discursive aspects highlighted by Jesus (2018) proved to be crucial in the identification of neological units present in the media and social networks.

Our preliminary findings show that *coronavírus* is regarded as a synonym of SARS-CoV-2. Like in other languages, this unit was the object of some graphic instability (*corona*, *corona virus*); it is frequently mentioned as a new type of coronavirus (*novo coronovírus*); it is used as the first element of compound words (*coronaditadura*, *corona-histeria*, *coronatroika*), also found in the context of loanwords (*coronabond*, *corona room*, *corona-app*); the use of prefixes as a time marker (*pre-*, *post-*) occurs often (*pré-coronavírus*, *pós-coronavírus*), while suffixes like *-phobia*, *-phobic* convey the fear of the pandemic (*coronafobia*, *coronafóbico*).

As *coronavírus*, *COVID-19* is regularly used as a synonym of SARS-CoV-2. Similarly, we have observed several processes of variation, such as graphic variation (*COVID-19*, *Covid-19*, *covid*) and gender assignment (*a/o covid*). Phonological neologisms, related to wordplays or puns, have been found mainly in the social networks: *covidizer* ("que ouvi dizer", 'that I heard'), *covidiano* ("quotidiano", 'daily'), *covidar* ("convidar", 'to invite') or the reflexive *covidar-se* (COVID+*infetar-se*, 'become infected'). The initialisms *a.C.* and d.C. (*antes/depois da COVID-19*) are associated with

an era, being equivalents to before/after Christ (*antes/depois de Cristo*). The phonological adaptation of loanwords resulted in neologisms as *covidiota*. Prefixation (*anti-*, *euro-, pre-, post-*) and suffixation (*-ade*, *-ário, -eiro, -ês, -iano, -ico*) processes were also highly productive.

Pandemia is frequently associated with prefixes (*pre-*, *post-*) to delimit a period, and other units related to the military semantic field (*combate à pandemia*, 'fighting the pandemic'; *lutar contra a pandemia*, 'to fight against the pandemic'; *batalha contra a pandemia*, 'battle against the pandemic'), showing that war metaphors concerning the COVID-19 pandemic also occurs in Portuguese. Phonological neologisms regarding wordplays or puns are frequent as well: *fraudemia* ('fraudemic'), *infodemia* ('infodemic'), *pãodemia* ('breademic', a recipe whose designation conveys the idea of the homemade bread trend during the pandemic).

Phrasal noun constructions stood out in all the neologism candidates under study. The unit categorized as a noun (*coronavírus*, *COVID-19*, *pandemia*) emerged simultaneously as an adjective: noun+adjective (*geração corona/pandemia*, *língua covid*). Additionally, the structure noun+preposition+noun (*tempos de corona/pandemia*, *pandemia/surto/vítima de covid-19*) was often recurrent, admitting alternative structures with articles, and other prepositions.

The prefix *tele-* conveys the idea of distance and forms words related to the use of telephones or television. However, in this new context, we observed that *tele-* expressed generally not only the concept of distance but also physical absence from the workplace or other events. The physical distance imposed by the pandemic was extended to the online world (internet and other telecommunication means), therefore, *tele-* is not necessarily related to television as before when speaking about *teletrabalho* ('teleworking'), *telescola* ('teleschool'), or *telepraxe* ('telehazing').

We believe that this research demonstrates the vitality of lexical neology processes from our synchronic lexicographic material in the domain of *COVID-19* in a specific period (December 2019–July 2021). Additionally to contributing to the neology field, this work will also result in the collection of detailed synchronic lexicographic material from the European Portuguese variety. Only the future will tell whether the creative linguistic phenomenon that emerged from the pandemic will persist in the Portuguese language (namely the loss of the neologism status of particular units while being incorporated in the current language lexicon) or whether it will be a source of occasionalisms circumscribed in time and space while the COVID-19 outbreak lasts.

Bibliography

a) Monographs, edited volumes and articles in edited volumes or journals

Alves, Ieda Maria (1990): *Neologismo: criação lexical*. São Paulo: Ática.
Alves, Ieda Maria/Maroneze, Bruno (2018): Neologia: histórico e perspectivas. In: *GTLex* 4(1),5–32.
Boulanger, Jean-Claude (1979): Néologie et terminologie. In: *Néologie en Marche* 4, 9–116.
Boulanger, Jean-Claude (2010): Sur l'existence des concepts de «néologie» et de «néologisme». Propos sur un paradoxe lexical et historique. In: *Atas do I Congrés Internacional de Neologia de les Llengues Romàniques*, 31–73.
Bueno Ruiz, Pedro J. (2020): When neologisms don't reach the dictionary: occasionalisms in Spanish. In: En Gavrillidou, Zoe; Mitsiaki, Maria; Filatouras, Asimakis (eds.): *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion*. Alexandroupolis: Democritus University of Thrace, 333–341.
Cabré, M. Teresa (1993): *La terminologia. Teoría, metodologia, aplicaciones*. Barcelona: Antártida/Empúries.
Cabré, M. Teresa/Freixa, Judit/Solé, Elisabet (2002) (ed.): *Lèxic i neologia*. Barcelona: Universitat Pompeu Fabra.
Correia, Margarita (1998): Neologia e terminologia. In: Mateus, Maria Helena & Correia, Margarita (coord.). *Terminologia: questões teóricas, métodos e projectos*. Mem Martins: Publicações Europa-América, 59–74.
Correia, Margarita/Lemos, Lúcia San Payo de (2005): *Inovação lexical em português*. Lisbon: Edições Colibri & Associação de Professores de Português.
Correia, Margarita et al. (2006): O Observatório de Neologia do Português Europeu – ONP: criação e apresentação. In *Atas do XX Encontro da Associação Portuguesa de Linguística*, 471–482.
Correia, Margarita (2009): *Os dicionários portugueses*. Lisbon: Editorial Caminho.
Cunha, Celso/Cintra, Lindley (1984): *Nova gramática do Português Contemporâneo*. Porto: Figueirinhas.
Dal, Georgette/Namer, Fiammetta (2016): À propos des occasionnalismes. In: V*Congrès Mondial de Linguistique Française*, 1–18.
Dressler, Wolfgang U. (1981): General principles of poetic license in word formation. In: Geckeler, Horst; Schlieben-Lange, Britigitte; Trabant, Jürgen; Weydt, Harald (eds.): *Logos semantikos. Studia linguistica in honorem Eugenio Coseriu* 1921–1981, vol. 2, Berlin: Mouton de Gruyter, 423–431.
Freitas, Tiago/Ramilo, Maria Celeste; Arim, Eva (2010): Os neologismos nos meios de comunicação social portugueses. In: Oliveira, Miguel, Jr. (org.): *Estudos de corpora. Da teoria à prática*. Lisbon: Edições Colibri & ILTEC, 129–144.
Guerrero Ramos, Gloria (2017): Nuevas orientaciones de la terminología y de la neología en el ámbito de la semántica léxica. In: *Rilce: Revista de Filología Hispánica* 33(3),1385–1415.
Guilbert, Louis (1975): *La créativité lexicale*. Paris: Libraire Larousse.
Jesus, Ana Maria Ribeiro de (2018): Tipologias dos neologismos: breve percurso histórico. In: *GTLex* 4(1),54–67.
Lino, Teresa (1988): Banco de neologismos do português contemporâneo – balanço de uma experiência. In *Atas do IV Encontro da Associação Portuguesa de Linguística*, 177–194.
Lino, Teresa (2003): Terminologia e indústria das línguas. In: *Atas do VII Simpósio Ibero-Americano de Terminologia*, 25–39.

Lino, Teresa (2018): Portuguese lexicography in the era of Internet. In: Fuertes-Olivera, Pedro (ed.): *The Routledge Handbook of Lexicography*, London, New York: Routledge, 608–618.
Lino, Teresa (2019): Neologia e neonímia em língua portuguesa: critérios de identificação. In: *Linha D'Água*, 32(3),9–23.
Pruvost, Jean/Sablayrolles, Jean-François (2003): *Les néologisms*. Paris: Presses Universitaires de France.
Rey, Alain (1976): Néologisme. Un pseudo-concept? In: *Cahiers de Lexicologie* 28, 3–17.
Rio-Torto, Graça (2020): Renovação do léxico no Português brasileiro e europeu. Da neologia técnico-científica à neologia expressiva, humorística, lúdica. In: *LaborHistórico* 6(3),196–223.
Sablayrolles, Jean-François (1997): Néologismes: une typologie des typologies. In: *Cahiers du CIEL 1996–1997*, 11–48.
Sablayrolles, Jean-François (2006): La néologie aujourd'hui. In: Gruaz, Claude (dir.): *À la recherche du mot: de la langue au discours*, Limoges: Lambert-Lucas, 141–157.
Sablayrolles, Jean-François (2016): Créativité lexicale et dictionnaristique. PhD seminar at NOVA University of Lisbon, March 3.
Thorne, Tony (2020): Spotlight on COVID: Pandemic language and the role of linguists. London: King's College London. [https://www.kcl.ac.uk/news/spotlight-on-covid-pandemic-language-and-the-role-of-linguists-1, last access: July 28, 2021].

b) Dictionaries, glossaries, and corpora

CETEMPúblico = *Corpus de Extractos de Textos Electrónicos MCT/Público*, Lisboa: Linguateca. [https://www.linguateca.pt/CETEMPublico/, last access: July 31, 2021].
Dicionário Caldas Aulete. Rio de Janeiro: Lexikon Editora Digital. [https://www.aulete.com.br/, last access: August 10, 2021].
Dicionário de Termos Médicos. Porto: Porto Editora, 2003–2021. [https://www.infopedia.pt/dicionarios/termos-medicos, last access: August 10, 2021].
Dicionário Moderno da Língua Portuguesa. Porto: Porto Editora, 2008.
DILP = *Dicionário Infopédia da Língua Portuguesa*. Porto: Porto Editora, 2003–2021. [https://www.infopedia.pt/, last access: August 10, 2021].
DPAH = *Dicionário do Português Atual Houaiss*. Lisboa: Círculo de Leitores, 1st edn. 2011.
DPE = *Dicionário de Português para Estrangeiros*. Porto: Porto Editora. [https://www.infopedia.pt/, last access: August 10, 2021].
DPLP = *Dicionário Priberam da Língua Portuguesa*. Lisboa: Priberam, 2008–2021. [https://dicionario.priberam.org/, last access: August 10, 2021].
Glossary *A covid-19 na língua*. In: Ciberdúvidas da Língua Portuguesa [https://ciberduvidas.iscte-iul.pt/artigos/rubricas/idioma/covid-19-na-lingua/4059, last access: August 16, 2021].

Ieda Maria Alves, Beatriz Curti-Contessoto, Lucimara Costa

COVID-19 terminology and its dissemination to a non-specialised public in Brazil

1 Introduction

The COVID-19 pandemic has impacted numerous sectors at different levels and has imposed a radical change in the pace of life in societies across the globe. Its consequences can be seen in various social spheres: health systems have been on the brink of collapse, the economies of many nations have been almost paralyzed, and people have adopted other ways of social contact, principally virtually. In all these areas, language has been present. Inevitably, it has also been influenced itself by this pandemic context.

A partially technical vocabulary related to COVID-19 quickly became part of everyday life, introduced mainly by news and official bodies. Daily bulletins with data on the numbers infected, cured and dead, contagion and vaccination maps, information about hygiene care, use of personal protective equipment, social behaviour rules, curfews and possible treatments have been daily transmitted to a large part of the world's population.

In Brazil, more specifically, in addition to this partially technical vocabulary, it has been possible to observe the recurrence of lexical units closely related to political and economic issues. Some of these reflect the stance of denial by the federal government in the face of the pandemic, and others concern the attitudes of the Brazilian population regarding the ways in which they dealt with restrictions on social interaction.

In order to describe the characteristics of the terminology being disseminated in Brazil, the project *Study and dissemination of COVID-19 terminology* was proposed, which is being developed under the coordination of Professor Ieda Maria Alves with the support of the Institute of Advanced Studies (IEA) at the University of São Paulo. The dissemination of this terminology will be achieved through a dictionary that is under development. This work is aimed at a Brazilian public, which,

Ieda Maria Alves, Dep. de Letras Clássicas e Vernáculas Av. Prof. Luciano Gualberto, 403 – Sala 04 – 2° andar. Cidade Universitária São Paulo – SP / Brasil, CEP: 05508–010,
e-mail: iemalves@usp.br
Beatriz Curti-Contessoto, Dep. de Letras Clássicas e Vernáculas Av. Prof. Luciano Gualberto, 403 – Sala 04 – 2° andar. Cidade University of São Paulo/FAPESP/ Brasil, CEP: 05508–010,
e-mail: bfcurti@gmail.com
Lucimara Costa, Dep. de Letras Clássicas e Vernáculas Av. Prof. Luciano Gualberto, 403 – Sala 04 – 2° andar. Cidade University of São Paulo/FAPESP/ Brasil, CEP: 05508–010,
e-mail: lucimara@usp.br

https://doi.org/10.1515/9783110798081-011

in general, has difficulty in understanding this specialised language and, often, has problems interpreting the guidelines and information in official health documents regarding the prevention, transmission, diagnosis and treatment of the disease.

Thus, as a contribution towards alleviating the numerous problems caused by the pandemic, it is hoped that, through an examination of the syntax and lexicon in Portuguese, the present study may serve to minimise the effects caused by the misunderstanding of medical language, contributing to the accessibility and divulgation of scientific knowledge by disseminating pandemic terminology to a non-medical audience, on an online platform. The study is also a demonstration of how the Human Sciences, especially the Language Sciences, can contribute to alleviating the effects of this pandemic in Brazil.

Thus, within the scope of this project, the study reported here aims to detect, analyse and discuss the characteristics of COVID-19 terminology, in particular the role of the adjective *novo* [new] in this terminology, the high recurrence of terms in the plural and the resemantization of some of the terminological units used. The present paper also discusses how these characteristics influenced the choices that have guided the creation of the proposed dictionary. This paper presents, therefore, the results of the analyses of these aspects, starting with a discussion of the relation between terminology and neology and arriving at the characteristic aspects of the macrostructural and microstructural choices about which some considerations were made.

2 The constitution of the corpora in the study

The present study was based on two corpora. The first of these is the *Official Corpus* (OC), which is composed of 993 texts concerning COVID-19 published on the following official websites: *Organização Mundial da Saúde* (OMS), *Organização Pan-Americana da Saúde* (OPAS), *Ministério da Saúde do Brasil* (MinSaude), *Agência Nacional de Vigilância Sanitária* (Anvisa), *Instituto Butantan*,[1] *Fundação Oswaldo Cruz* (Fiocruz)[2] and *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP).[3] Given the target public of the dictionary in preparation, the *Journalistic Corpus* (JC) was also created

1 The Butantan Institute is the main immunobiological producer in Brazil and is responsible for a large percentage of the production of hyperimmune serums and a large volume of the national production of vaccine antigens, which make up the vaccines used in the PNI (National Immunization Program) of the Brazilian Ministry of Health (Instituto Butantan 2021).

2 Created in 1900 as a pioneering initiative in the country, the Oswaldo Cruz Institute (OCI/Fiocruz) [. . .] constitutes a complex that generates knowledge, products and services in the biomedical area to meet the health needs of the Brazilian population (FioCruz 2021).

3 São Paulo Research Foundation is one of the main agencies for promoting scientific and technological research in the country (Agência FAPESP 2021).

with the purpose of selecting the most used terms in large circulation press vehicles in the Brazilian territory. This corpus, which is complementary, contains 460 texts collected from the following websites: *Folha de S. Paulo* (FSP), O *Estado de S. Paulo* (ESP) and *O Globo* (GLO).

These corpora were compiled following a methodology based on the web as corpus (cf. Kilgarriff 2013). For that, the tools BootCat Bootstrap Corpora and Terms from the Web, version 1.21 (Zanchetta et al. 2011), and AntConc (Anthony 2012) were used. First, the BootCat served to find the texts available until March 2021 on the web regarding the topic in question. Once the corpora were constituted, the compiled texts were treated through the AntConc program. In this program, the observation and selection of the lexical units present in the corpora were carried out. To do that, different types of lists (keywords, wordlist, clusters and concordance) were created.

These resources were used to identify terminological candidates present in the corpora. These candidates were then checked in the light of terminological assumptions. Among the criteria used in this process, there are those presented by Barros (2007), which were used in order to verify the degree of lexicalization of terminological phrases and to determine the limits of the syntagmatic terminological units.

3 Main characteristics related to the constitution of COVID-19 terminology

Rey (1995: 11) states that the history of science and technology has shown that the relations between terminology and neology can be found since the first people began to name concepts and elements of their environment. The author stresses this character of the constitution of terminologies by emphasising that "terminology is fundamentally concerned with names and the process of naming" (Rey 1995: 11).

COVID-19, an infectious disease, is caused by the new coronavirus, which is a virus of the "family of Coronaviridae that causes infections in humans and animals (*e.g.*, respiratory diseases, gastroenteritis *etc.*)"[4] (*Houaiss* 2012, online, our translation). Since the disease is caused by a previously unknown virus, which is nonetheless part of a family of existing and already known viruses, it has been named *novo coronavírus* [new coronavirus] by the World Health Organization (WHO). This designation, released on 30 December, 2019 by the Director-General of the WHO, Tedros Adhanom, specifies that, notwithstanding the fact that it is a virus of an existing family, this new member has its own characteristics, which correspond to the first

4 Original: "família dos coronavirídeos, causadores de infecções em seres humanos e em animais (p. ex., doenças respiratórias, gastrenterite *etc.*)".

meaning attributed to the adjective *novo* [new]: "1 that which was born or appeared recently, which has little lifetime, little time of existence (it is said esp. of living beings)"[5] (*Houaiss* 2012, our translation).

These new features of *novo coronavírus* have determined, in the studied terminology, the creation of several syntagmatic formations made up of the adjective *novo* and its gender and number inflections in Brazilian Portuguese (*novos*, *nova (s)*), which attribute to these lexical units the characteristics expressed in the aforesaid meaning. Among these formations, the present paper mentions the most frequent, mainly used in the plural because they refer, in various contexts, to a large number of *novas cepas* [new strains], *novas linhagens* [new lines] and *novas variantes* [new variants] of the virus. Some examples[6] extracted from OC and JC are presented below:

(1) Os vírus mudam sempre, e são essas mudanças que levam ao surgimento de *novas cepas* ou variantes, que podem ou não ser mais perigosas. Mais de quatro mil mutações foram descritas no Sars-CoV-2 desde o início da pandemia. (<JC_GLO_211220>)

(2) A cada mudança do vírus, são geradas também *novas linhagens*, o que justifica a necessidade de novos estudos para uma melhor compreensão dos fatores clínicos e epidemiológicos relacionados à doença. (<OC_FIOCRUZ_110221>).

(3) Há o risco de que *novas variantes* do vírus possam acabar "driblando" a vacina e muitos especialistas estimam que com o tempo, seja necessário atualizar a vacina e reaplicá-la, como ocorre com as vacinas da gripe. (<JC_ESP_170121>)

Louis Guilbert, in two pioneering works on the constitution of terminologies, emphasises the role of classifier or specifier of adjectives, which, employed after the substantival core of a syntagma, allow the integration of this syntagma into another specialty area. In *La formation du vocabulaire de l'aviation* (1965a), Guilbert highlights the importance of the adjective *aérien* in the formation of aviation vocabulary, which is related to a new type of transport:

5 Original: "1 que nasceu ou apareceu recentemente, que tem pouco tempo de vida, de existência (diz-se esp. de seres vivos)".

6 Our translations: (1) Viruses are always changing, and it is these changes that lead to the emergence of *new strains* or variants, which may or may not be more dangerous. More than four thousand mutations have been described in Sars-CoV-2 since the beginning of the pandemic. (2) With each change in the virus, *new strains* are also generated, which justifies the need for further studies to better understand the clinical and epidemiological factors related to the disease. (3) There is a risk that *new variants* of the virus could end up "bypassing" the vaccine, and many experts estimate that, over time, it will be necessary to update the vaccine and reapply it, as with flu vaccines.

> When the transfer from an old semantic field to a new semantic field takes place in the form of an integrating syntagma, the second element, which has an adjectival form, is the main linguistic instrument of this transfer. The most common among transfer adjectives is *aérien*.[7]
>
> (Guilbert 1965a: 198, our translation)

Some examples of formations with *aérien* cited by Guilbert (1965a) are: *argonaute aérien*, *navigateurs aériens, route aérienne*, *voiture aérienne*, *voyageurs aériens*.

In another work, *Le vocabulaire de l'astronautique* (1965b), Guilbert studies astronautics terminology, which was born in the early 1960s. Unlike his study on aviation, which has a diachronic character, this new study refers to a synchrony from 1961 to 1963. This work mentions some syntagmatic terms, whose high frequency is noted and which are formed with the adjectives *cosmique* and *spatial*. This terminology includes, respectively, *biologie cosmique, firmament cosmique* and *cabine spatiale*, *plate-forme spatiale*, among others. In these examples, it is observed that the "key adjectives" that indicate that a term is integrated (or is becoming integrated) into astronautics terminology are *cosmique* and *spatial*, which are part of several syntagmatic terms in this specialised area.

Humbley, in his study entitled *La néologie terminologique* (2018), refers to e-commerce terminology (*commerce électronique*). He affirms that it originated from an ancestral domain, or source domain (*domaine source*), which is the domain of commerce. In this example, the element that enables the transfer between domains is the adjective *éléctronique*.

Regarding COVID-19 terminology, it is possible to highlight the use of the adjective *novo*, which is quite recurrent, which enables us to consider it as a characteristic of this terminology related to the novelty of the pandemic. However, what is also found is that, in the case of *novo coronavírus*, the adjective *novo* loses its qualifier character, as occurs in *novas cepas*, *novas linhagens* and *novas variantes*, and starts to occupy the role of classifier / typifier, given the fact that what is being formed is a proper name from an already existing proper name (cf. Neves 2018). The term *coronavírus*, which is part of this formation, was already a name given to a disease through the name of a virus.

In addition to the terminological characteristic regarding the use of the adjective *novo* in several terminological syntagmas found in OC and JC, these corpora exposed another interesting recurrence: the constancy of the plural form in several substantive terms. This use of the plural occurs because the effects of infection caused by the new coronavirus are not singular. Indeed, they can be exacerbated by different diseases, including diabetes, hypertension, heart and lung diseases.

---

7 No original: "Quand le transfert d'un champ sémantique ancien à un champ sémantique nouveau se réalise sous forme de syntagme d'intégration, le second élément de forme adjectivale est l'instrument linguistique principal du transfert. Le plus fréquent parmi les adjectifs de transfert est *aérien*.".

The set of these diseases is designated by *comorbidades* [comorbidities], as exemplified below:

(4) A probabilidade de uma pessoa obesa desenvolver a forma grave da Covid-19 é alta independentemente da idade, do sexo, da etnia e da existência de *comorbidades* como diabetes, hipertensão, doença cardíaca ou pulmonar.[8] (<OC_FAPESP_030920>)

Consequently, the forms of prevention and treatment of the disease are also multiple, as are the possible side-effects of the vaccine, and this characteristic is expressed by terms used primarily in their plural form. This use is exemplified by the terms *eventos adversos* [adverse events] and *equipamentos de proteção individual* (EPIs) [personal protective equipment (PPE)].

Eventos adversos [adverse events] related to the COVID-19 vaccine are varied, with headache, fever, myalgia, diarrhoea, nausea and localised pain being the most common:

(5) Até o dia 04/02/2021, foram notificados ao Ministério da Saúde 7.768 *eventos adversos* supostamente associados às vacinas contra a Covid-19. Desses, 7.686 foram classificados como eventos não graves e 82 foram classificados como graves. Os eventos mais comuns foram cefaleia, febre, mialgia, diarreia, náusea e dor localizada.[9] (<OC_MinSaude_210720>)

Equipamentos de proteção individual (*EPIs*), including surgical masks, respirators and glasses, protect professionals who work with health equipment and patients with infectious diseases:

(6) Usadas em conjunto com outros *equipamentos de proteção individual (EPIs)*, como máscaras cirúrgicas, respiradores e/ou óculos, aumentam a proteção oferecida aos profissionais que estão atuando nos equipamentos de saúde.[10] (<OC_FAPESP_060520>)

8 Our translation: (4) The probability of an obese person developing the severe form of Covid-19 is high regardless of age, gender, ethnicity and of the existence of *comorbidities* such as diabetes, hypertension, heart or lung disease.

9 Our translation: (10) Until 4 February, 2021, 7,768 *adverse events* allegedly associated with Covid-19 vaccines were reported to the Ministry of Health. Among them, 7,686 were classified as non-serious events and 82 were classified as severe. The most common events were headache, fever, myalgia, diarrhoea, nausea and localised pain.

10 Our translation: (11) Used in conjunction with other *personal protective equipment (PPE)*, such as surgical masks, respirators and/or glasses, they increase the protection offered to professionals who are working with health equipment.

Another characteristic of some terms related to COVID-19 concerns the resemantization process that these units went through. Among these terms, this study mentions, as examples, *confinamento* [confinement], *lockdown* and *quarentena* [quarantine]. These lexical units were reinterpreted through the extension of their concepts, which began to be used specifically in relation to the pandemic.

By way of illustration, the following excerpts[11] include the terms *confinamento*, *lockdown* and *quarentena* highlighted in bold:

(7) Se não tiver política econômica, monetária e fiscal, vamos caminhar para uma depressão. O coronavírus não é uma gripe comum, tem muitas diferenças em relação às crises econômicas típicas do livro-texto. É um choque de oferta e, em seguida, de demanda, com *lockdown (confinamento)*. (<JC_GLO_210320>)

(8) Todos os pacientes infectados com o novo coronavírus precisam ser hospitalizados e colocados em *quarentena*, a menos que a condição seja grave, é melhor evitar o ambiente hospitalar; para isolar em casa, é melhor abrir a janela para ventilação. (<OC_FAPESP_040320>).

Confinamento refers to the "act or effect of imposing (by the authority, the government) a determined residence on an individual, away from social contact", or to "prison isolation"[12] (*Houaiss* 2012, our translation). In its turn, *lockdown*, a borrowing from English that had already occurred in Brazilian Portuguese, designates the "action of isolating people, confining them for a certain time (at home, on a ship, in a hospital, *etc*.), for safety reasons (during pandemics, for example)"[13] (*Houaiss* 2012, our translation) and competes with its translated form, *confinamento*, within the scope of COVID-19. However, these two terminological units were reinterpreted in the pandemic context. Although the definition of *lockdown* has one meaning that refers to a type of security measure adopted in pandemics in general, both units came to designate, in a more specific way, measures that had to be established by the government for the purpose of containing the new coronavirus in society.

Another terminological unit that reveals this extension of meaning is *quarentena*. It is true that its concept already included semantic features related to: "40-day

---

11 Our translations: (7) If there is no economic, monetary and fiscal policy, we are heading for a depression. Coronavirus is not a common flu, it has many differences from typical textbook economic crises. It is a clash between supply, followed by demand, and *lockdown (confinement)*. (8) All patients infected with the new coronavirus need to be hospitalized and placed in *quarantine*; unless the condition is severe, it is best to avoid the hospital environment; to isolate at home, it is better to open the window for ventilation.

12 Original: "ato ou efeito de impor (a autoridade, o governo) uma residência determinada a um indivíduo, longe do contato social", ou a um "isolamento prisional".

13 Original: "ação de isolar pessoas, confinando-as por certo tempo (em casa, num navio, num hospital *etc*.), por medida de segurança (durante pandemias, p. ex.)".

period", "set of measures and restrictions that specifically consisted in the isolation, for a certain time (origin. 42 days), of individuals and goods from regions where epidemics of contagious diseases were raging", "set of restrictions and/or isolation, for variable periods of time, imposed on individuals or loads from countries where epidemics of contagious diseases occur" and "Lent",[14] among others (*Houaiss* 2012, our translation). In none of its meanings, however, is there a close relation with the context of the COVID-19 pandemic. Therefore, what this term reveals as a characteristic marked by the pandemic context is precisely the fact that it refers to one of the government measures to contain this disease that tried to close cities, which forced the population to isolate themselves in their homes and whose duration varied according to the occupancy rate of hospital beds. This terminological unit also concerns the isolation during the incubation period of the new coronavirus practiced by people who have had possible contact with patients or have travelled through regions and situations of high risk of contagion (through air travel, for example), as well as by patients, so that they do not spread the virus.

4 Characteristics of COVID-19 terminology and terminographical implications

The corpus relating to the terminology of COVID-19, presented above, was created as a result of the elaboration of a terminological dictionary aimed at non-specialised speakers in the medical field and with little formal education. The aspects of this terminology that are emphasised in the present paper – the use of the adjective *novo*, the recurrence of the plural form in various substantive and syntagmatic terms, the process of resemantization that various lexical units have gone through – guided some aspects of the constitution of the dictionary that are highlighted in this section.

The motivation for the creation of this dictionary arises from the fact that Brazil is a country of gigantic dimensions, with a population of 213,797,113 million inhabitants on 11 January, 2021, according to data from the Brazilian Institute of Geography and Statistics (IBGE). This population is highly heterogeneous with regard to its educational level, which varies according to several factors, especially geographical and social.

Data from the Functional Literacy Indicator (INAF), released in 2018, indicate that 29% of the Brazilian population has difficulty in interpreting texts and performing

14 Original: "período de 40 dias", "hist.med conjunto de medidas e restrições que consistia esp. no isolamento, durante certo tempo (orign. 42 dias), de indivíduos e mercadorias provenientes de regiões onde grassavam epidemias de doenças contagiosas", "infect conjunto de restrições e/ou isolamento, por períodos de tempo variáveis, impostos a indivíduos ou cargas procedentes de países em que ocorrem epidemias de doenças contagiosas" e "Quaresma".

simple mathematical operations in their daily activities. The INAF tests are applied to Brazilians who are between 15 and 64 years old, with the aim of analysing their skills and practices in reading, writing and mathematics aimed at everyday life. According to the INAF results, Brazil has 29% of functional illiterates. A functionally illiterate person is a literate individual "whose poor literacy leads him / her to write very poorly and not be able to interpret what he / she reads"[15] (*Houaiss* 2012, our translation). The INAF divides the functionally illiterate into two groups: *absolute* (8%), who cannot read words or phrases and telephone numbers, for example, and *rudimentary* (21%), who have difficulty in identifying ironies and sarcasm in short texts, and in performing simple operations such as calculating money (INAF 2021).

Based on these data and the interpretation difficulties manifested by about 29% of Brazilians between 15 and 64 years old, the macro- and microstructure of the dictionary were designed aimed at Brazilian speakers with different levels of education, especially those with little schooling.

Regarding the macrostructure, the terms will be presented according to the number of occurrences verified in the studied corpora. As commented in the previous section, these corpora indicate the predominant usage in a plural form in several substantive terms and in several syntagmatic terms, since the effects of the infection caused by the new coronavirus can reach several organs, cause different diseases and require different forms of treatment. Consequently, the forms of prevention and treatment of the disease are also multiple, expressed by terms used primarily in the plural form. This usage was exemplified with the terms *comorbidades*, *eventos adversos* and *equipamentos de proteção individual* (*EPIs*). The dictionary will highlight the plurality expressed by these terms by presenting their respective entries in the plural form. It is important to say that these terms do not correspond to the concept of *pluralia tantum*, as they also occur in the singular, much less frequently, within the scope of the studied terminology. In COVID-19 terminology, they are most commonly used in the plural because they refer to the multiple impacts of the virus.

Another characteristic that influenced the selection of syntagmatic terms that make up the dictionary's nomenclature concerns the formations in which the adjective *novo* appears. It is noted that this adjective attributes not only the quality of being new, but, above all, it brings a specificity to the concept designated by the simple term with which it is associated. In this case, the syntagmatic term *novo coronavírus* is an example. Thus, *novo coronavírus*, which, as explained above, designates another type of virus, different from the other coronaviruses that make up the family of Coronaviridae, appears as a terminological entry.

15 Original: "cuja alfabetização precária o leva a escrever muito mal e a não conseguir interpretar o que lê".

The presentation of the terms and their respective entries will follow an onomasiological organization, according to the categories established by the *DeCS/MeSH* (Health Sciences Descriptors 2021), released by the Pan American Health Organization (PAHO). According to the classification established by this organization, the terms will be separated into categories, such as: causative agent, anatomy, prevention, diagnosis, disease, treatment, equipment, among others.

As the dictionary is intended for a broad and non-specialised public in the medical field, it will follow the principles of an international tendency towards using simplified language, more easily understood by users who are not specialised in the area in question. Due to these principles, the definitions and explanatory or complementary notes, in relation to the term, are being written in plain language, which designates texts understandable by different types of speakers.

Named *plain language* in English, this type of communication, in Brazil, has been designated *linguagem simples* [simple language], *linguagem clara* [clear language], *linguagem cidadã* [citizen language], *acessibilidade textual e terminológica*[16] [textual and terminological accessibility, *inteligibilidade*[17] [intelligibility], but the first term is predominant in the country: "a communication is in plain language if its wording, structure, and design are so clear that the intended audience can easily find what they need, understand what they find, and use that information", according to the Plain Language Network (Plain Language 2021), an international association for plain language supporters and practitioners around the world, which includes members from 30 countries and at least 15 languages.

In Brazil, some government initiatives are adopting and disseminating this type of language. The *Rede Linguagem Simples* [Simple Language Network], for example, was created by the federal government (*Empresa Brasil de Comunicação – EBC*), which is "a space for debate and fostering the construction of initiatives that promote the use of simple language"[18] (Agência Brasil 2021, our translation).

In 2016, in the State of São Paulo, the state government launched a manual called *Orientações para adoção de linguagem clara* [Guidelines for adopting clear language], which

> should be understood as a guide for the elaboration of a *Clear Language* (CL), to be used in the way the São Paulo State government disseminates its information on the Internet, especially with regard to the meaning of technical expressions routinely used by specialists of the various areas of government action, in order to make them more accessible to the understanding of the common citizen.[19]
>
> (Governo do Estado de São Paulo 2016, our translation)

16 cf. Cortina Silva et al. (2021).

17 cf. Carvalho and Rebechi (2021).

18 Original: "um espaço de debate e de fomento para construção de iniciativas que promovam o uso da linguagem simples".

19 Original: "deve ser entendido como um roteiro para elaboração de uma *Linguagem Clara* (*LC*), para ser acoplada à maneira como o governo do Estado de São Paulo divulga suas informações na

In 2019, the City Hall of São Paulo launched the *Programa Municipal de Linguagem Simples* [Municipal Simple Language Program], with the release of a booklet, called *Princípios de uma Linguagem Cidadã e Manual de boas práticas de redação da Carta de Serviços da Prefeitura de São Paulo de Linguagem Cidadã* [Principles of a Citizen's Language and Manual of good practices for writing the Charter of Services of the City Hall of São Paulo for Citizen's Language], in order to

> help PMSP servants to write texts for the population in a clear, inclusive and understandable way for people of all genders, classes and educational levels, discarding the use of bureaucratic and formal language used in public offices, which is also often used to address São Paulo society.[20] (Prefeitura de São Paulo 2021)

The concept of *simple language* does not imply the use of simplistic or informal language or the elimination of information elements. It seeks to use an understandable language with clear information, avoiding problems frequently reported by readers due to the use of long sentences, passive verbs, acronyms, abbreviations, little used adjectives and unexplained terms. Preference should be given to common words, better known by users and to the usual syntactic order of the language (Cortina Silva *et al.* 2021). Taking into account the precepts of simple language, the definitions in the dictionary are being written, preferably, with one syntactic period, since, in most cases, it is possible to elaborate a definition consisting of a single sentence.

Each of these definitions uses, primarily, the classic binary categorization *close genre + specific differences*. The close genre has the function of initial descriptor of the definitions and rescues the conceptual content of its hyperonym (a more generic term in relation to other terms) and, therefore, of the general characteristics of the term, thus expressing the general category or class to which this element belongs. The specific differences present the particularities that distinguish the term from others of the same class.

Other definitions, such as *extensional* ones, in which elements that characterise the term are enumerated, are also being used in cases where it is necessary to enumerate several features. An example of this type of definition corresponds to the term *medidas preventivas*: *medidas preventivas consistem em lavar as mãos, usar máscara, usar álcool em gel, evitar aglomerações* [preventive measures consist of washing hands, using a mask, using alcohol gel, avoiding agglomerations].

Internet, em especial no que diz respeito ao significado de expressões técnicas utilizadas rotineiramente por especialistas das diversas áreas de atuação governamental, de modo a torná-las mais acessíveis à compreensão do cidadão comum".

20 Original: "auxiliar servidoras e servidores da PMSP a redigirem textos destinados à população de forma clara, inclusiva e compreensível a pessoas de todos os gêneros, classes e níveis de instrução, descartando o uso da linguagem burocrática e formal utilizada nas repartições públicas, e que muitas vezes também é usada para se direcionar à sociedade paulistana".

By way of illustration, two examples of definitions are presented, accompanied by their respective entries (with English equivalent, context of use, explanatory note). These terms were extracted from the causative agent category (*carga viral*)[21] and from the diagnostic category (*comorbidades*):[22]

> *Carga viral* s.f. Quantidade de vírus encontrada em amostras de sangue ou em outros fluidos da pessoa infectada.
> Ing. *viral load*
>
> Uma *carga viral* mais elevada foi observada mais frequentemente nos pacientes do sexo masculino e nos mais idosos. Febre e atralgia (dor nas articulações) foram os sintomas mais associados a uma *carga viral* elevada. (<CO_FAPESP_290421>)
>
> Nota: O termo *fluido* designa uma substância que corre como um líquido, a exemplo de sangue.
>
> *Comorbidades* s.f.pl.
>
> Duas ou mais doenças presentes em uma pessoa, como diabetes, doença cardíaca ou pulmonar.
> Ing. *comorbidities*
>
> Identificar grupos de maior risco para adoecimento, agravamento e óbito: Idosos; Pessoas com *comorbidades*: Diabetes, HAS, Doenças cardíacas/cerebrovasculares, DPOC, Renal, Obesidade, Câncer, Transplantados, Anemia Falciforme (<CO_MinSaude_ 271120>)
>
> Nota: O termo *comorbidade* é usualmente empregado no plural porque se refere a pelo menos duas doenças.

The examples presented above show, respectively, *carga viral* and *comorbidades* as terminological entries. In *carga viral*, an explanatory note was presented to elucidate the meaning of the term *fluido* [fluid], as it is little known to the general public.

The fact that the term *comorbidades* was registered in its plural form reveals one of the aspects observed in relation to the constitution of COVID-19 terminology, which was explored in the previous section. This aspect is not only reflected in the expression of this term, but also in its definition, since the idea of plurality, as a

---

21 Our translation: *Viral load* f.n. Amount of virus found in blood samples or other fluids from the infected person. Eng. *viral load*. A higher *viral load* was seen more frequently in male and older patients. Fever and arthralgia (joint pain) were the symptoms most associated with a high *viral load*. (<OC_FAPESP_290421>). Note: The term fluid designates a substance that flows like a liquid, such as blood.

22 Our translation: *Comorbidities* f.n. in pl. Two or more diseases present in a person, such as diabetes, heart disease or lung disease. Eng. *comorbidities*. To identify groups at higher risk for illness, aggravation and death: Elderly; People with comorbidities: Diabetes, SAH, Heart/Cerebrovascular Diseases, COPD, Kidney disease, Obesity, Cancer, Transplants, Sickle Cell Anemia. (<OC_MinSaude_ 271120>). Note: The term *comorbidity* is usually used in the plural because it refers to at least two diseases.

set, is incorporated within it. The preference of using it in the plural, instead of its singular form, is explained by the note in this entry.

5 Final considerations

The aim of the present study was to detect, analyse and discuss the characteristics of COVID-19 terminology, in particular the role of the adjective *novo* in this terminology, the high recurrence of terms in the plural and the resemantization of some of the terminological units used. It also sought to fulfil the objective of specifying how these terminological characteristics are reflected in the constitution of a COVID-19 dictionary, which is under preparation.

Regarding the use of the adjective *novo*, which proved to be quite recurrent in this terminology, it was found that, when added to the term *coronavírus*, this element assumes the function of being a classifier – which does not generally occur with this particular adjective, especially when placed before the noun that it determines. In other terminologies, such as those studied by Guilbert and Humbley, mentioned before, classifier adjectives have another nature, less common, and are directly related to a specialty domain that is formative.

The observation of corpora also revealed a high productivity of lexical units employed in their plural form. Through the examples mentioned in the present paper, it is noted that this plurality builds the concepts designated by the studied terms, as they mark, in their expression, which comorbidities are in focus in the pandemic context in which we live, the side effects associated with the COVID-19 vaccine and the personal protective equipment (PPE) necessary for professionals on the front lines of the fight against the pandemic. In these cases, the use of the plural indicates a very particular characteristic of the COVID-19 terminology: that the idea of a set, of the collective, is essential to the concepts to which these units refer, in addition to the fact of relating, from a semantic-conceptual point of view, these terms to the pandemic.

It was also found that some terms went through a process of resemantization, thanks to the fact that lexical units like *confinamento*, *lockdown* and *quarentena*, among others, had their meaning expanded to reflect conceptual information related to the context of the pandemic.

These characteristics influenced the choices that guided the creation of the proposed dictionary, which is a terminological dictionary aimed at non-specialised readers in the medical field with little formal education.

Bibliography

Agência Brasil (2021): *Rede quer facilitar linguagem de serviços à população*. [https://agenciabrasil.ebc.com.br/geral/noticia/2021-03/rede-quer-facilitar-linguagem-de-servicos-populacao, last access: 18 August, 2021].

Agência FAPESP (2021): *São Caetano do Sul investe na atenção primária para enfrentar a pandemia*. [https://agencia.fapesp.br/sao-caetano-do-sul-investe-na-atencao-primaria-para-enfrentar-a-pandemia/33604/, last access: 10 August, 2021].

ANVISA (2021): *Agência Nacional de Vigilância Sanitária*. [https://www.gov.br/anvisa/pt-br>, last access: 10 August, 2021].

Anthony, Laurence (2012): *AntConc (Version 3.5.8) [Windows]*. Tokyo, Japan: Waseda University. [http://www.laurenceanthony.net/software/antconc/, last access: 14 July, 2020].

Barros, Lídia Almeida (2007): *Conhecimentos de Terminologia geral para a prática tradutória*. São José do Rio Preto, SP: NovaGraf.

Carvalho, Yiuli S./Rebechi, Rozane (2021): Inteligibilidade e convencionalidade em textos de divulgação da área médica em português brasileiro. In: *Rev. Estud. Ling.*, Belo Horizonte, 29(2), 959–998.

Cortina Silva, Asafi F./Delgado, Heloísa O. K./Finatto, Maria J. B. (2021): Acessibilidade textual e terminológica para o português brasileiro: pesquisa, estratégias e orientações de [re]escrita. In: *Revista Moara*, 58, 322–343

FioCruz (2021): *Fundação Oswaldo Cruz*. [https://portal.fiocruz.br, last access: 20 September 2021].

Governo do Estado de São Paulo (2016): *Orientações para adoção de linguagem clara*. [http://www.governoaberto.sp.gov.br/wp-content/uploads/2017/12/orientacoes_para_adocao_linguagem_clara_ptBR.pdf, last access: 20 September 2021].

Guilbert, Louis (1965a): *La formation du vocabulaire de l'aviation* (1861–1891). Paris: Larousse.

Guilbert, Louis (1965b): *Le vocabulaire de l'astronautique*. Paris: Publications de l'Université de Rouen.

Health Sciences Descriptors (2021): DeCS – *Descritores em Ciência da Saúde*. [https://decs.bvsalud.org/, last access: 20 August, 2021].

Houaiss, Antônio (2012): *Grande dicionário Houaiss*. Rio de Janeiro: Instituto Antônio Houaiss. [https://houaiss.uol.com.br/corporativo/apps/uol_www/v5-4/html/index.php#4, last access: 8 November, 2021].

Humbley, John (2018): *La néologie terminologique*. Limoges: Lambert Lucas.

IBGE (2021): *Instituto Brasileiro de Geografia e Estatística*. [https://www.ibge.gov.br, last access: 30 October, 2021].

Instituto Butantan (2021): *A serviço da vida*. [https://butantan.gov.br, last access: 20 October, 2021].

INAF (2021): *Indicador de Alfabetismo Funcional*. [https://alfabetismofuncional.org.br, last access: 14 October, 2021].

Journal Folha de S. Paulo (2021): *Folha*. [https://www.folha.uol.com.br, last access: 14 October, 2021].

Journal O Estado de S. Paulo (2021): *Estadão*. [https://www.estadao.com.br, last access 14 October, 2021].

Journal O Globo (2021): *O Globo*. [https://oglobo.globo.com, last access: 10 September, 2021].

Kilgarriff, Adam/Rigau, Irene (2013): EsTenTen, a vast web corpus of Peninsular and American Spanish. In: *International Conference on Corpus Linguistics (CILC2013*). Alicante, Spain, 12–19. DOI: https://doi.org/10.1016/j.sbspro.2013.10.617.

PAHO (2021): *Pan American Health Organization*. [https://www.paho.org/en, last access: 15 September, 2021].

Plain Language (2021): *What is plain language?* [https://plainlanguagenetwork.org/plain-language/what-is-plain-language/, last access: 14 August, 2021].

Prefeitura de São Paulo (2021): *Cartilha Princípios de uma linguagem cidadã e manual de boas práticas de redação da Carta de Serviços da Prefeitura de São Paulo*. [https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/chamadas/3_colocado_grupo_1539290157.pdf, last access: 14 October, 2021].

Ministério da Saúde (2021): *Ministério da Saúde*. [https://https://www.gov.br/saude/pt-br, last access: 10 September, 2021].

Neves, Maria Helena de Moura. (2018): *A gramática do português revelada em textos*. 1ed. São Paulo: Ed. Unesp.

Rey, Alain (1995): *La terminologie. Noms et notions*. Paris: PUF.

Tcacenco, Lucas/Rodrigues da Silva, Bruna/Finatto,Maria J. B. (2020): Acessibilidade textual e terminológica. In: Revista *GTLex*, 3(2), p. 197–224.

WHO (2021): World Health Organization. [https://www.who.int, last access: 20 October, 2021].

Zanchetta, Eros/Baroni, Marco/Bernardini, Silvia (2011): Corpora for the masses: the BootCaT front-end. *Proceedings of the Corpus Linguistics Conference 2011*. Birmingham: University of Birmingham.

Rute Costa, Margarida Ramos, Ana Salgado, Sara Carvalho,
Bruno Almeida, Raquel Silva

Neoterm or neologism? A closer look at the determinologisation process

1 Introduction

This paper arises within the current communication urgency experienced throughout the pandemic. From its onset, several new lexical units have permeated the overall media discourse, as well as social media and other channels. These units convey information to the public regarding the 'severe acute respiratory syndrome' namely *COVID-19*.[1] In addition to its worldwide impact healthwise, the pandemic generates noteworthy influence in the linguistic landscape, and as a result, a significant number of neologisms have emerged. Within the scope of our ongoing research, we identify the neologisms in European Portuguese that are related to the term *COVID-19* via form or meaning. However, not all the new lexical units identified in our corpus containing *COVID-19* in its formation can unequivocally be regarded as neoterms (terminological neologisms). Accordingly, this article aims not only to reflect on the distinction between neologism and neoterm but also to explore the determinologisation process that several of these new lexical units experience.

Following the introduction, this paper is divided into 9 sections. In section 2, we begin by making a brief theoretical reflection concerning neological processes

1 In this paper, the term *COVID-19* is the preferred form referring to 'acute respiratory syndrome'.

Rute Costa, Margarida Ramos, NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Campus de Campolide, Colégio Almada Negreiros, 1099-032 Lisboa, Portugal
Ana Salgado, NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Campus de Campolide, Colégio Almada Negreiros, 1099-032 Lisboa; ACL, Academia das Ciências de Lisboa, Instituto de Lexicologia e Lexicografia da Língua Portuguesa, Rua da Academia das Ciências de Lisboa, 19, 1249-122 Lisboa, Portugal
Sara Carvalho, CLLC, Centro de Línguas, Literaturas e Culturas da Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal; NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Campus de Campolide, Colégio Almada Negreiros, 1099-032 Lisboa, Portugal
Bruno Almeida, ROSSIO Infrastructure, Colégio Almada Negreiros, Campus de Campolide da NOVA, 1099-085 Lisboa, Portugal; NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Campus de Campolide, Colégio Almada Negreiros, 1099-032 Lisboa, Portugal
Raquel Silva, VOH.CoLAB – Value for Health CoLAB, Nova Medical School, Edifício CEDOC I, Rua do Instituto Bacteriológico, n.os 5, 5-A e 5-B, 1150-190 Lisboa, Portugal; NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Campus de Campolide, Colégio Almada Negreiros, 1099-032 Lisboa, Portugal

https://doi.org/10.1515/9783110798081-012

and determinologisation. Then, in section 3, we describe the method used to compile the corpus, the *CoronaCorpus*, which unfolds in 2 sub-corpora: the Press*CoronaCorpus* and the LSP*CoronaCorpus*. The Press*CoronaCorpus* is composed of texts published in the Portuguese media between November 2020 and July 2021. This corpus has been processed via Sketch Engine,[2] with the purpose of identifying neological lexical constructions occurring in non-specialised communication related to the emergence of the pandemic. Among such neological constructions, both neologisms and neoterms were identified. The latter are defined as terms that are "specifically coined for a given general concept" (ISO 1087:2019, §3.4.12). The second sub-corpus, the LSP*CoronaCorpus*, is composed of official documents produced by healthcare agencies, professionals and scientists. In the context of this research, this corpus plays the role of a reference corpus.

In section 4, our corpus is explored by means of simple and advanced queries for extracting the spelling variants of *COVID-19*. Section 5, on the other hand, is focussed on the *COVID-19* acronym, its behaviour in discourse and the re-categorisation of *covid-* as a formative in Portuguese. In section 6, we then proceed with the analysis of morphosyntactic and semantic formation of the neologisms and neoterms identified in the Press*CoronaCorpus*, to better grasp the process underpinning neology, in what concerns both form and meaning. Furthermore, this section aims to describe some of the behaviours depicted by the new elements containing the form *COVID-19* and which occur in non-specialised communication contexts. The migration of terms from specialised to non-specialised contexts points towards a shift in status, from term to non-term. Such change, in some cases, results from determinologisation processes which are analysed further.

Next, in section 6, we analyse the lexical units and terms found in our corpus and describe the respective neological and determinologisation processes.

In section 7, we focus on the lexicographic treatment of four neologisms that have been registered in Portuguese e-dictionaries available in Portugal, namely the *Dicionário Priberam da Língua Portuguesa* (DPLP)[3] and the *Dicionário da Língua Portuguesa* of Porto Editora (DLP).[4]

Finally, based on our corpus analysis workflow, as well as on the systematic comparison of the aforementioned dictionary entries, a template for a lexicographic article targeted at neologisms is put forward in section 8, illustrated by the entry *covid*. This proposal aims, on the one hand, to address the detected inconsistencies in lexicographic representation in the cited Portuguese resources and, on the other hand, to respond to this form's behaviour in the corpus, namely as an element used to create new words (e.g. *covid* + *-ário*).

2 https://www.sketchengine.eu/ (last access: 10 June 2022).
3 https://dicionario.priberam.org/ (last access: 10 June 2022).
4 https://www.infopedia.pt/dicionarios/lingua-portuguesa (last access: 10 June 2022).

Overall, the inclusion of neoterms in dictionaries entails several challenges, such as their morphosyntactic classifications, their definition, and which domain label they should or should not be assigned.

2 Neological processes and determinologisation

2.1 Neological processes

In the case of a pandemic, there is an enhancement of neological processes, which emerge more or less spontaneously to quickly resolve communication issues associated with scientific phenomena, which go beyond the understanding of the non-specialised public. New words resulting from these processes are considered to be neologisms. Within non-specialised communication, neologisms may arise from the need to have a communicative impact on the overall community when referring to previously non-existent realities and may even stem from highly specialised contexts. On the other hand, there are neoterms, which designate new specialised concepts produced in a given domain of knowledge. Contrary to a neologism, which is formed spontaneously in relation to communication issues, a neoterm is often formed consciously to designate a concept and distinguish it from others in the concept system to which it belongs, so that it can be used in specialised discourse with a low degree of ambiguity.

Both neologisms and neoterms are linguistic phenomena that are morphosyntactically manifested via the creation of new lexical units, or, semantically, via the attribution of new meanings to already existing lexical units. Neologisms can, therefore, be analysed according to different perspectives. In line with what has been stated by Lino, "neologisms are simultaneously a manifestation of the evolution of a language and the evolution of knowledge, both of which happening at an extremely quick pace" (2019: 10). Terminologists and lexicologists look at the phenomenon of neology from a different standpoint. In terminology science, a neologism is defined as a "term that is specifically coined for a given general concept" (ISO 1087:2019, §3.4.12), whereas in general language, a neologism is defined as "a new word" or "a new meaning of an existing word in the language" (Pruvost and Sablayrolles 2003). The difference is quite significant. That is probably the reason why, in terminology science, the terms "neoterm", "terminological neologism" (ISO 1087:2019) and "neonymy" (Rondeau 1984) have been created to differentiate the conceptual level from the linguistic one. In lexicology and lexical morphology, neologisms are mainly studied as part of word formation and semantics, the latter further exploring topics related to semantic shifting or semantic extension.

Relevant linguistic phenomena include, among others, the formation of terms, the study of collocations and phraseologies, lexical and semantic relations, formal

and semantic neology, as well as variation. In several of these phenomena, the identified linguistic change often has an impact on the dictionary's macro- and microstructure, as well as on the lexical units to be selected to feed the lexicographic resource. These lexicographic activities require that corpora are consistently maintained and up-to-date to detect neologisms and neoterms in time to meet the users' demands.

2.2 Determinologisation processes

Determinologisation (Guilbert 1975, Galisson 1978, Meyer and Mackintosh 2000) is the process by which a term is transformed into a general language word or expression. In these cases, the term does not refer to a concept anymore and, therefore, it is no longer part of a concept system within a given domain. Hence, there is a semantic or conceptual shift prompted by the elimination of one or more essential characteristics of the concept, thereby leading the term to lose its identity and specificity.

Nová (2018: 387) goes further and considers that determinologisation corresponds to the process by which "a scientific term, during its way from a field specialist to a layperson, loses its accuracy, gets new connotations, and the word can be even moved to refer to a completely different thing".

Semantic shift and term variation are the main axes for the study of the specialised lexicon appearing in scientific, technological and technical texts and discourses, both written and oral. Therefore, linguistic change in form and meaning is a dynamic phenomenon that cuts across the entire lexicon. The time-lapse during which meaning is formed, from point A to point B, is recorded in dictionaries, encyclopaedias, vocabularies and ontologies through the choice of lemmas (lexical unit or term) and the definition of the concept and/or the explanation of its meaning.

In terminology, the definition stabilises the relationship between the lexical unit (form) and the specialised concept from a domain of knowledge, in a given period and cultural, political or social context. Meaning is thus fixed in time, which could be short in areas such as science and technology. This finding, resulting from long years of studying the lexicon, allows us to introduce the concept of 'short diachrony', which can be observed in real time. Short diachrony occurs when one observes linguistic change at the level of lexical units, mostly specialised, because of immediate changes in knowledge structures, e.g., when a new concept is introduced in a specialised domain, which has a direct impact on the lexicon. This linguistic change is identified, analysed and classified but it must also be registered, described or defined, and dated in dictionaries to preserve linguistic heritage.

The 'short diachrony' observed in scientific, technical and technological texts contrasts with the 'long diachrony' typically studied in historical linguistics. Short diachrony is of great importance for the construction and update of corpora. In

science and technology, corpora age very quickly when it comes to the study of the lexicon, requiring a constant renewal of the texts that constitute them and constant observance of the published literature in the domains under study. Naturally, there are specialised domains in which more change is observed than in others, with varying rhythms and dimensions.

This is why it is relevant that the Press*CoronaCorpus* corresponds to a monitor corpus, because as stated by Sinclair (1996): "It became clear some years ago that the assumption of a finite limit on a corpus for any length of time was an unnecessary restriction."[5] As depicted in the following sections, it is possible to observe, in a relatively short time span, the appearance and disappearance of lexical units, as well as variation phenomena.

3 Press*CoronaCorpus*: the corpus of analysis

This work has been carried out via the analysis of a dedicated monolingual [EU Portuguese] corpus comprising both a journalistic and a Language for Special Purposes (LSP) subcorpora. As referred to in the introduction, the journalistic corpus – Press*CoronaCorpus* – was compiled using Sketch Engine, namely the WebBootCaT technology, along with a manual identification of texts publicly available on the internet, to capture newspapers and magazines related to COVID-19 topics. On the other hand, as mentioned above, we also retained official documents produced by healthcare agencies, healthcare professionals and scientists, making up the LSP*CoronaCorpus*. As such, and for the purposes of our study, the journalistic corpus is the corpus of analysis, whereas the LSP-based is the reference corpus. The latter is used to verify semantic shifts, differences in neologism formation, as well as differences in usage, if needed.

The spectrum of Press*CoronaCorpus* is 9 months wide. It is a dynamic corpus intended to represent a snapshot of the language between November 2020 and July 2021 – a time window within the pandemic context – to observe how covid-driven new forms, and corresponding spelling variants, productively entered the Portuguese lexicon.

The texts were gathered during this period on a weekly and sometimes biweekly basis, resulting in a large collection of different text types. This collection was organised according to the activities and targeted audience of the texts, culminating in a text typology as systematised in Table 1:

5 http://www.ilc.cnr.it/EAGLES96/corpustyp/node19.html #SECTION00090000000000000000 (last access: 10 June 2022).

Table 1: Text typology according to social activities and targeted audience.

Text type	Newspaper	Magazine
Generic	503	
Business & Economics	150	36
Sports	36	2
Environment & Garden		3
Fashion & Socialite		14
Health & Lifestyle		7
IT & Electronics		8
Travel & Culture		24

The number of texts is significantly larger for newspapers when compared with magazines. On the other hand, the most salient type is Generic, whose percentage is 73% when compared with Business & Economics and Sports, whereas the collection of magazines is mostly related to Business & Economics, with 38%.

The capture of media texts on the Internet is not a straightforward task. Due to their increasing subscription-based model, some media pages are not fully available; consequently, the texts were not collected in a balanced quantity throughout the period we have set. To overcome this drawback, we decided to (i) store the collected texts in .txt format, (ii) organise them by trimesters, and (iii) attribute a descriptor to each of them (e.g. PT-NP-GE-2020-11 – which stands for a Portuguese generic newspaper published in November 2020). Such a decision is tied with the manual task of corpus metadata annotation, a process that took place during the compilation of the collected texts with Sketch Engine.

With regard to part-of-speech annotation, we resorted to a tagger embedded in Sketch Engine, specifically the Portuguese FreeLing tagset, since the texts that build up Press*CoronaCorpus* are in EU Portuguese. In short, by merging those two types of corpus annotations, we developed an annotated corpus enriched with metadata, text type and corresponding topic.

Regarding the overall metrics of the corpus of analysis, despite the number of texts not being quantitatively balanced throughout the trimesters, the metrics are considerably robust for the diachronic spectrum under study. As seen in Table 2, the corpus has a little over 40 million tokens and more than 30 million words.[6]

6 "A word is a type of token. Words are tokens which begin with a letter of the alphabet" (https://www.sketchengine.eu/guide/glossary/?letter=W) (last access: 10 June 2022).

Table 2: The metrics of the corpus (tokens, words and number of texts).

		PressCoronaCorpus			
		15nov2020-15jan2021	**16jan-15mar2021**	**16mar-15mai2021**	**16mai-11jul 2021**
Tokens	**41 610 865**	11 163 278	17 052 708	9 557 012	3 837 867
Words	**32 930 803**	8 849 548	13 438 800	7 640 911	3 001 544
Texts	**986**	278	405	214	89

4 Exploring the corpus of analysis: the example of COVID-19

For the corpus exploration, we resorted to both simple and advanced queries, depending on the results we were aiming at. Whereas the former use special characters, such as /*/ (e.g. *covid** – from which we obtained matches like *covid*; *covid-19*; *covid19*; and so forth), the latter resorted to Corpus Query Language (CQL), with regular expressions (REGEX) at its core. For instance, to capture *covid* as a monolexical unit without punctuation and digits, we resorted to the following CQL: [word = "*covid*"][!(word = "-.*|.*19")]. The same text mining strategy was used for *Covid* and *COVID*, but with different regexes for the first element [word =], given the different letter cases.

The focus of this paper is the term *COVID-19*, which will be further explored in section 5.

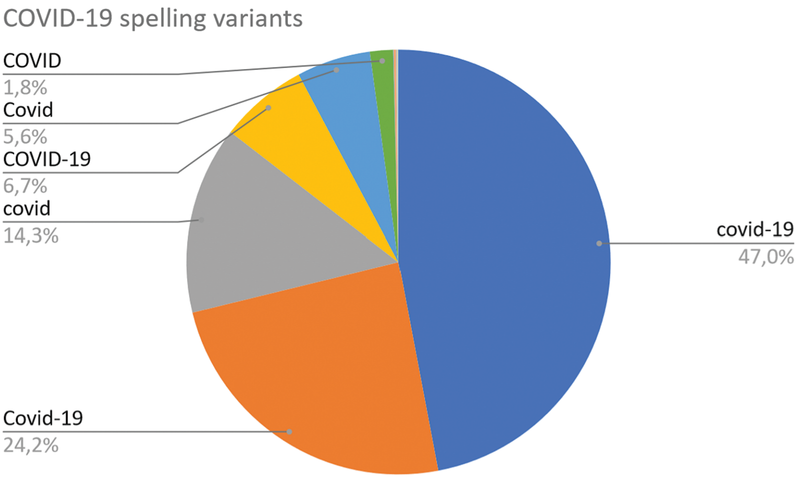
Table 3: *COVID-19* spelling variants and frequencies for Portuguese language.

covid-19	**10611**
Covid-19	5455
covid	3224
COVID-19	1521
Covid	1274
COVID	399
covid 19	37
Covid 19	24
COVID19	11

Table 3 (continued)

covid-19	**10611**
covid19	8
Covid19	6

Table 3 depicts several spelling variants found in the corpus, with *covid-19* as the most frequent. Interestingly, despite being the form officially used in texts written by experts (much like OMS for EU Portuguese and WHO for English), *COVID-19* has a minor representation (6,7%) in the corpus when compared to *covid-19* (47%), as represented in Graph 1.

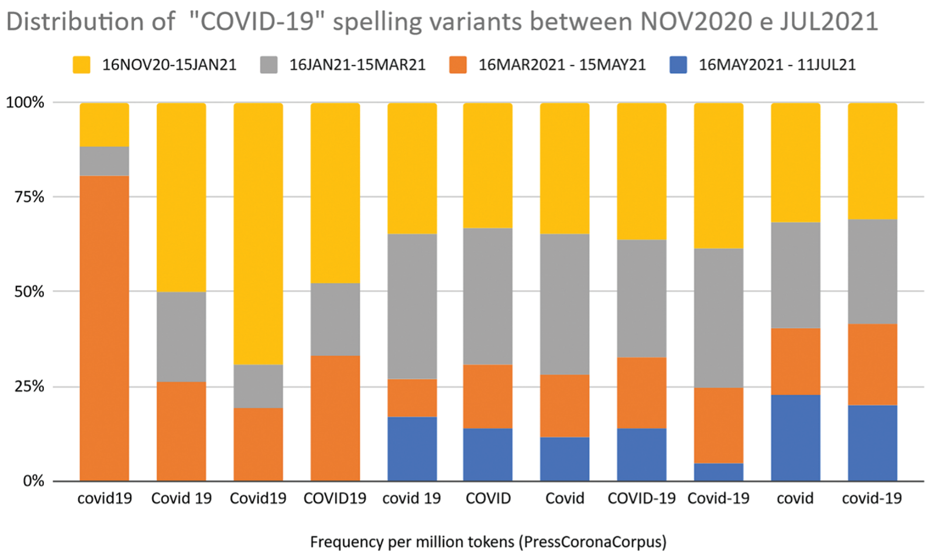


Graph 1: *Covid-19* vs *COVID-19*.

The remaining spelling variants, in turn, not only have a low number of occurrences, but also a short period of evidence throughout time. This can be observed in the diachronic spectrum of the corpus, as represented in Graph 2.

Graph 2 depicts the diachronic distribution of the *COVID-19* spelling variants between November 2020 and July 2021. The spelling variants are systematised by trimester and according to per million frequencies, like the example for *covid-19* in Table 4.

Thus, and as observed in the corpus, some spelling variants of *COVID-19* seem to have a shorter lifespan than others. This is the case of variants that have no evidence in the last trimester (16MAY21-11JUL21). These short-life variants, namely *covid19*, *Covid 19*, *Covid19*, and *COVID19* may reflect the instability of these new lexical units in discourse, given its recent lexicalisation.

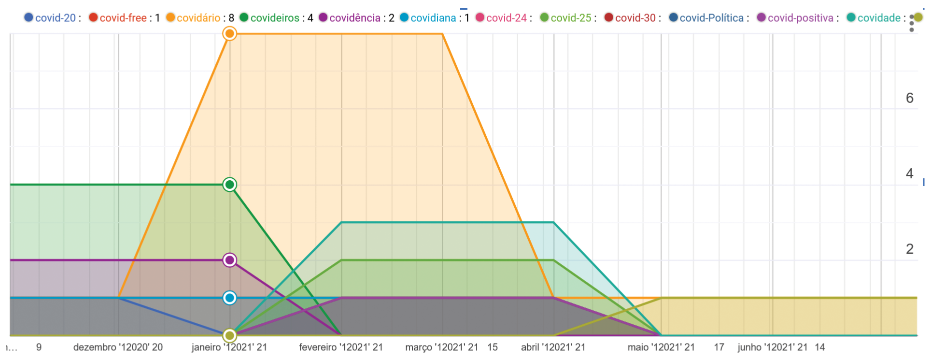


Graph 2: Diachronic distribution of *COVID-19* spelling variants (Nov. 2020 – Jul. 2021).

Table 4: Systematisation of the spelling variant *covid-19* according to the per million frequencies by trimester (Nov. 2020 – Jul. 2021).

{covid}*	**16NOV20–15JAN21**	**16JAN21–15MAR21**	**16MAR2021–15MAY21**	**16MAY202 – 11JUL21**
covid-19	296.24 per million tokens • 0.03%	296.24 per million tokens • 0.03%	207.07 per million tokens • 0.021%	192.03 per million tokens • 0.019%

To further reflect this short diachrony topic, Graph 3 illustrates the time span of 13 forms starting with *covid-* that we have identified in the corpus.



Graph 3: Some examples of lexicalisations with short diachrony/short time span.

Focusing on January 2021 (highlighted with bullets in Graph 3), we can observe that the lexical unit *covidário* has its first occurrence in the corpus in December 2020, and quickly reached its highest frequency in January, with [8] occurrences. This number of occurrences remains until March, finally dropping to [1] occurrence in April and remaining as such until July 2021. On the other hand, the lexical units *covideiros* and *covidência*, with [4] and [2] occurrences respectively, do not appear further beyond February – an evidence over which we hypothesise a short time span, given the diachronic spectrum of the corpus. The lexical units *covidade* [3], *covid-25* [2] and *covid-positiva* [1] are 3 examples with a short time span, which begins in February and ends in April. Finally, the lexical unit *covidiana* [1] is the form denoting the longest time span among the forms under focus here, i.e., it occurs between November 2020 and May 2021. It should be noted that November is not the first attestation of the form, but corresponds to the onset of the corpus compilation.

5 COVID-19: general considerations on word formation

The first word to be analysed is *COVID-19*. *COVID-19* is an acronym formed by the initialisms of the constituents of the polylexical term ***co**rona**vi**rus **d**isease 20**19***. The acronym results from a truncation process: *co-* and *vi-*, truncated elements, plus the initial *d* from disease. The number *19* points to 2019, the year when the World Health Organization (WHO) first learned about this new virus on 31 December 2019, following a report of a cluster of cases of an unknown pneumonia disease in Wuhan, People's Republic of China.[7] The WHO's proposal complies with the best practices recommended in 2015 by this organisation for the designation of new human infectious diseases.[8] The English acronym has been imported into Portuguese, despite its complete lack of connection to the corresponding Portuguese polylexical term, which is *doença do coronavírus 2019*. *COVID-19* is, therefore, a hybrid acronym formed by the initials of the constituents of the English polylexical unit and the reduction of the 2019 number, which becomes stabilised in speech and language through a lexicalisation process. The speaker assimilates the acronym as a monolexical unit and integrates it into his/her lexicon functioning as a noun, as attested in our corpus:

---

7 https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19 (last access: 10 June 2022).
8 https://www.who.int/publications/i/item/WHO-HSE-FOS-15.1 (last access: 10 June 2022).

(1) ***a*** *COVID-19 veio complicar ainda mais a adequada gestão das doenças raras* [COVID-19 **(f.n.)** further complicated the adequate management of rare diseases]

(2) *A presença física no escritório nunca foi obrigatória para a generalidade dos nossos colaboradores ao longo dos últimos oito meses de coexistência com* ***a*** *COVID-19* [Physical presence at the office was never mandatory for most of our employees over the last eight months of coexisting with COVID-19 **(f.n.)**]

(3) *Será possível exigir testes de diagnóstico para* ***a*** *COVID-19 nos estabelecimentos de saúde* [The requirement of diagnostic tests for COVID-19 **(f.n.)** will become possible in health facilities]

Given that *COVID-19* refers to a disease (*doença*, in Portuguese), the term's gender in Portuguese is feminine and behaves mostly like a noun in discourse. Nevertheless, the use of the term in the masculine gender also occurs, as documented in the DPLP (cf. Figure 1) ("É também usado como substantivo masculino." [It is also used as a masculine noun.] (DPLP, 2021), since the speakers, by metonymy, designate the disease after the virus causing it, which also happened with other diseases (e.g., the Zika virus and the Zika disease):

COVID-19

(acrónimo do inglês *coronavirus disease 2019*, doença de coronavírus 2019 [ano em que a doença foi identificada pela primeira vez])

nome feminino

[Medicina] Doença infecciosa respiratória, causada pelo coronavírus SARS-CoV-2, cujos sintomas podem incluir febre, tosse, dificuldades respiratórias e cansaço, e que, em alguns casos, pode progredir para pneumonia ou falha respiratória.

Nota: Também se escreve com minúsculas (*covid-19*). É também usado como substantivo masculino.

Figure 1: Lexicographic article *COVID-19* in DPLP.

Our research has identified several new lexical units integrating the initial acronym, whereby *covid-* takes on the role of a formative, thus undergoing a re-categorisation. According to Bauer (2003: 330) formative is "a recurrent element of form which correlates with derivational behaviour in some way and yet cannot be identified with a morph".



Figure 2: The evolution of the disease designation until it appears as a formative.

As represented in Figure 2, *COVID-19* is the acronym of the polylexical unit 'coronavirus disease 2019' which in turn, through an ellipse process, loses the hyphen and the reference to the year. When the acronym *COVID* is written in lowercase (*covid*), the lexical unit is perceived as a noun. The term *covid-19*, a feminine noun, is the most frequent in the corpus, with *covid* being the most productive form from a morphological standpoint. Hence, *COVID* is a highly productive lexical unit, behaving as a base form upon which a set of morphological processes act, giving rise to new lexical units while undergoing re-categorisation processes.

The lexical categories of the term *covid* found in Press*CoronaCorpus* are noun (1) and adjective (2):

(1) O futuro, se a **covid** nos permitir, é sempre risonho, nós temos de olhar para a frente sempre com perspetivas de construção [The future, if covid allows us, is always bright, we must always look forward with constructive perspectives].

(2) Ontem, numa corrida só com quatro atletas – por precauções **'covidianas'**, quando nos 5000 marcha, durante muitos minutos, andaram todos juntos [Yesterday, in a race with only four athletes – for 'covidian' precautions, when in the 5000 march, for many minutes, they all walked together].

Regarding word formation, we have observed the typical word processes: derivation and composition. In the case of derivation, words are formed through the addition of a prefix or a suffix to the base, which can be a stem, a theme (stem + thematic vowel) or a lexical unit. In this paper, we focus on the acronym *COVID*, which is the base form of the lexical units we are analysing in section 6. The acronym is a lexical unit formed by the word formation process in which an initialism is read syllabically and is a morphological constituent on which word formation operations take place, thereby allowing the creation of formal and semantic derived monolexical neologisms.

Composition, in turn, is a word formation process that operates by concatenating two or more word stems or two or more words. In Portuguese, there are two types of composition: morphological composition, which concatenates stems according to the principles of morphological word formation, and morphosyntactic composition, according to which properties of syntactic structures and properties of morphological structures are combined. The examples we selected illustrate cases of morphosyntactic compounds: *ala covid* and *doente covid*. These compounds are formed by an adjunct structure, that is, they are constituted by two nouns, with similar behaviour to nominal syntactic structures. The right constituent (*covid* in both examples) functions as a nominal modifier, generating a new lexical unit.

Considering the dynamics of the acronym *covid* and since it is highly productive from a morphosyntactic point of view, we use the term to describe formation of lexical units in which *covid* is an endocentric formative.

6 Analysis of morphosyntactic and semantic formation of the neologisms and neoterms

In this section, we observe the occurrences of the base form *covid* to identify behaviours and regularities in word formation, as well as its associations with other elements (prefixes, suffixes), to verify lexical productivity and determine the semantic component of the elements. Based on the data analysis, we will justify whether these words can be considered neoterms or, on the contrary, if having a term in their formation corresponds to a 'false' intuition. As stated by Lombard/Huyghe/Gygax (2021), the neological intuition is an essential feature of neologisms, which can vary according to the individuals and the regularity of lexical creativity processes.

One of the most regular and productive word-formation processes in Portuguese is derivation. Derivation is distinguished from composition in that, contrary to the latter, there is only one autonomous unit of lexical meaning – the base – to which an affix (prefix or suffix) is added to form a new lexical unit. As examples of derived words, we selected the noun *covidário* [covidarium] (whose occurrence in the corpus remained stable over four months) and the adjective *covidiana* [covidian] (one of the first occurrences identified in the corpus). In these examples, two suffixes *-ário* and *-(i)ano* or *-(i)ana* (with a feminine inflexion mark) are added to the base *covid-*. These suffixes occurring after the base determine the lexical category of the newly formed nominal base derivatives (*-ário*, which forms nouns, and *-(i) ano*, which is highly productive in the formation of adjectives), and are also carriers of semantic values. *Covidário* denotes a place (i.e. it has a locative value, being a locative denominal noun according to Rio-Torto et al. (2013) where certain entities remain or are housed, such as in *aviário* (aviary), *berçário* (nursery), *fraldário* (diaper changing room), *infantário* (nursery school) and *solário* (solarium). In turn, the denominal adjective *covidiano* is formed by the adjectival suffix *-(i)ano*, which denotes a living being, as in *bacteriano* (bacterial).

Again, within the processes of derivation, *anticovid* is an example of a word formed by prefixation. The prefix *anti-* combines with the nominal base, *covid*, changing the base word's lexical category (*covid* N → *anticovid* ADJ). Due to its semantic, oppositional value, the prefix *anti-* is combined with bases denoting entities (a disease, in this case), without number inflexion.

As an example of morphosyntactic composition, we chose *covid-drive*, a hybrid compound that occurs twice in our corpus. In terms of its constitution, the noun *covid* is combined with another English-imported noun making up an [N + N] structure. Concerning morphosyntactic compounds, we highlight occurrences such as *ala covid* (covid ward) or *doente covid* (covid patient), which can be included within the so-called "modifying compounds" (Rio-Torto et al. 2013: 93), in that the second element modifies the noun. We are in the presence of an [N + N] structure:

(1) [[ala] N1 + [covid] N2] polylexical unit (compound)

(2) [[doente] N1 + [covid] N2] polylexical unit (compound)

In both examples (1) and (2), the noun *covid* qualifies the N1, allowing us to infer that the *ala covid* denotes a 'place where patients with the disease are housed/hospitalised'. In the second case, *doente covid* denotes a 'person with covid disease'. In these situations, time provides us with the answer of whether these compounds are going to become lexicalised or not and, consequently, if the lexicographer should describe them in a dictionary, which in these cases has already happened. Both *ala covid* and *doente covid* are neologisms, not because the process of word formation is innovative, but because the lexical distribution of their elements has a novelty effect.

From a semantic point of view, these neologisms are formed by lexical units belonging to the lexicon of general/current use, *ala* and *doente*, to which the term *covid* is associated. The question that arises is whether the compound resulting from the combination of these two units forms a neologism or a neoterm. Following our analysis, these units occur in specialised contexts, mostly in public health discourse, but they are not exactly terms because they do not belong to any particular specialised domain. This implies that, in the context of a lexicographic work, these units would not be classified as belonging to medicine, biology, virology, epidemiology or public health through domain labels. In these two examples, the specificity of the < covid > concept is nullified because the characteristics /Respiratory System Disorder/, /Pneumonia/ and /Viral Pneumonia/ are cancelled, thus losing the semantic value associated with specialised domains related to the disease.[9]

A different process is that of the lexical unit *covidário*, which is a derivative formed by analogy, for example, with *berçário* [nursery]. *Covidário* can be defined as an isolated space in a health facility dedicated to covid patients. In this case, the -ário suffix does not have a specialised sense and is therefore not a term, being classified as a neologism, although it occurs in specialised contexts, especially those related to hospitals. However, curiously, the DPL, for example, considers *covidário* as a term of Medicine, but does not consider *berçário* as a term belonging to that same domain, despite referring to hospitals in the definition. Have lexicographers been misled by the semantic value of the base *covid*? We find that whenever the formative *covid* appears, there is a tendency to consider the new lexical unit as a term [ex: *covidiota*, *covidivórcio*]

Lastly, the *covidiota* occurrence clearly results from a process of word formation in which the specialised sense of *covid* is lost. From our point of view, this is a process of determinologisation. *Covidiota* is divided into *covid* + *idiota*, being a morphological compound, which is used to refer derisively to a person who does not respect general

9 http://covidterm.imicams.ac.cn/#/search?isAdvanced=false&keyword=covid (last access: 10 June 2022).

safety measures, either voluntarily or involuntarily. This unit is a neologism of form and meaning. Formally, it is a portmanteau that corresponds to the blending of two lexical units in which one of the units is truncated: *id* being the end of the *covid* unit and the beginning of the *idiota* unit. We are faced with a haplology, which corresponds to the elimination of one of two consecutive syllables when they are identical or very similar (see Marquilhas 2014: 28). This neologism has no specialised value.

As we can see, it is not always self-evident whether a lexical unit is specialised or not. The fact that the *covid* acronym and its variants are morphologically productive and dynamic (see Table 5) requires an accurate analysis of each of the cases.

Table 5: The productivity of the *covid* acronym.

LEXICAL UNIT	POS	WORD FORMATION	WORD FORMATION
covid	N	Abbreviation of a polylexical unit	ACRONYM
covidário	N	[**covid**]*base* + [ário] *suf.*	DERIVATION
covidiana	ADJ	[**covid**]*base* +[(i)ana] *suf.*	DERIVATION
ala covid	N	[ala]*N* + [**covid**]*N*	COMPOSITION
covidade	N	[**cov**id]*base* + [**id**ade] *suf.*	DERIVATION [HAPLOLOGY]
covidiota	N	[**covid**]*base* + [**id**iota] N	COMPOSITION [PORTMANTEAU]

The determinologisation process is evident in the formation of new general language words, as shown by the examples in Table 6. Reusing the *covid* acronym as a formative element contributes to the process of determinologisation, since the core meaning is used in a superficial manner.

Table 6: Determinologisation process.

covidiota	[+ neologism] [– neoterm]	[composition | portmanteau]
covideiro	[+ neologism] [– neoterm]	[derivation | suffixation]
covidices	[+ neologism] [–neoterm]	[derivation | suffixation]
covidivórcios	[+ neologism] [–neoterm]	[composition | portmanteau]

These lexical units are used in a covid context, but they neither convey specialised features, nor do they belong to a conceptual system of a domain.

7 Lexicographic treatment of neologisms associated with COVID-19

After the extraction and analysis of neologisms from Press*CoronaCorpus*, we now move on to the lexicographic treatment of four neologisms – *covid*; *covidário*; *covidiana*; *anticovid* – previously selected and registered in Portuguese language e-dictionaries, namely DPLP and DLP.

These two lexicographic resources were selected because (i) they are available online, (ii) they are constantly updated with neologisms, both from general language and specialised language, and (iii) they have a very broad list of headwords, each having more than 100,000 entries.

DPLP is a contemporary Portuguese dictionary with about 133,000 lexical entries, whose headword list comprises general language vocabulary as well as terms from various specialised domains. This resource also offers the possibility of browsing entries in European Portuguese spelling, following the 1990 Portuguese Language Orthographic Agreement, and in Brazilian Portuguese, with and without the changes prescribed by this agreement.

DLP is a monolingual Portuguese dictionary that is integrated into the infopedia.pt service,[10] which provides 30 bilingual online dictionaries in several languages (Portuguese, Portuguese Sign Language, English, Spanish, French, German, Italian, Dutch, Chinese, Tetum and Greek). Following European Portuguese spelling, it has two versions: one according to the 1990 Orthographic Agreement of the Portuguese Language, and the other according to the previous standard, that is, the Portuguese-Brazilian Orthographic Agreement of 1945.

The pandemic, as the term itself implies (pan-, Greek *pan*, all), caused the rapid and simultaneous entry of new words – the aforementioned neologisms – in languages around the world. The urgency to publish neologisms of high daily frequency in real time and the need to satisfy the searches of dictionary users often lead to some rash decisions, not allowing lexicographers appropriate and timely reflection on the phenomena and consequent validation of data. Aware of this problem and of certain limitations, we proceed with the comparison of the lexicographic treatment of neologisms associated with the pandemic crisis, intending to answer the following questions:

1) Do DPLP and DLP register the neologisms detected in Press*CoronaCorpus* and which are currently under analysis?
2) Do DPLP and DLP account for these units as neologisms?
3) Do DPLP and DLP consider *covid* as a formative?

The first step was to check whether these units occur in DPLP and DLP. We found that the above-mentioned units are attested in both dictionaries.

10 https://www.infopedia.pt/ (last access: 10 June 2022).

As we can see in Figure 3, the neologisms *covidário* and *covidiano* show morphological structures specific to Portuguese, that is, they are considered words derived by suffixation (suffixes *-ário*; *-(i)ano*).



DPLP

co·vi·dá·ri·o
(*COVID[-19]* + *-ário*)
nome masculino
[Medicina] Local, devidamente isolado e equipado que, num estabelecimento de saúde, se destina ao atendimento e ao tratamento de doentes com suspeita ou confirmação de infecção por COVID-19 (ex.: *o hospital criou um covidário para reduzir os riscos de contágio*).

co·vi·di·a·no
(*COVID[-19]* + *-iano*)
adjectivo
Relativo a uma doença infecciosa respiratória causada pelo coronavírus SARS-CoV-2; relativo à COVID-19 (ex.: *confinamento covidiano*). = COVÍDICO

DLP

covidário
co.vi.dá.ri.o • kɔviˈdarju
nome masculino
MEDICINA área, em hospital ou outra instituição de saúde, especialmente preparada para o atendimento e tratamento de doentes com COVID-19, bem como de indivíduos suspeitos de estarem infetados pelo vírus que origina essa doença
De *COVID[-19]+-ário*

covidiano
co.vi.di.a.no • kɔviˈdjɐnu
adjetivo
relativo à COVID-19 (doença respiratória viral causada pelo coronavírus SARS-CoV-2)
De *COVID[-19]+-iano*

Figure 3: Lexicographic articles *covidário* and *covidiano* (DPLP, DLP).

The first lexicographic data analysed pertains to the formation of words. This information is shown in DPLP in italics between brackets, below the syllabic division of the words. DLP, on the other hand, makes use of an icon to show information about word formation, thereby requiring the user to hover the mouse cursor over the icon to access this information. The two lexicographic resources coincide in the analysis: *COVID-[19]* + *-ário*. The first point that catches our attention is the fact that this information does not indicate the possibility of *covid* being treated as a formative (*covid-*) for new words, indicating instead that these words – *covidário* and *covidiano* – are formed from the original acronym rather than the *covid* noun itself.

Another topic, often controversial in lexicography, lies in the use of the *Medicina* [Medicine] domain label in the *covidário* article in both dictionaries – especially when confronted with the word *covidiano*, which does not have any domain label. We may question whether this word belongs, in fact, to the medical domain or whether it constitutes a process of determinologisation. Even so, this topic goes beyond our scope, since domain labels in general language dictionaries, in many cases, only function as mere identifiers or word sense disambiguators. Moreover, the dictionaries do not inform us about the criteria for using this label, so we could only make assumptions regarding their use.

Moving on to the illustrative example of derivation by prefixation (starting with the prefix *anti-*), we have the *anticovid* entry. Although our corpus shows the

hyphenated spelling *anti-covid*, according to the Portuguese orthography, the word must be written in an agglutinated form, as evidenced by the entries in the dictionaries. However, while DPLP registers *anticovid*, DLP registers *anticovid-19*, where once again the word formation points to the *COVID-19* acronym (see Figure 4).


DPLP

an·ti·covid

(*anti-* + *COVID*)

adjectivo de dois géneros e de dois números

Que se destina a combater ou proteger contra a COVID-19 (ex.: *máscara anticovid; vacinas anticovid*).

DLP

anticovid-19

ɐ̃tikovid dəzɐˈnɔv(ə)

adjetivo invariável

que visa combater ou prevenir a COVID-19 (doença respiratória viral causada pelo coronavírus SARS-CoV-2)

De *anti*+*COVID-19*

Figure 4: Lexicographic article *anticovid* (DPLP, DLP).

We will now turn to the analysis of the units denoting the 'severe acute respiratory syndrome' *COVID-19* (see Figure 5). The orthographic forms are treated differently in each dictionary: the lemma registered in DPLP is the acronym in uppercase (*COVID-19*), while DLP chose as lemma the acronym together with its lowercase form, which can be considered as a spelling variant. The *COVID* entry is also present in DPLP, which further notes the lowercase form: "Também se escreve com minúsculas (covid)" [Also written in lowercase (covid)].


DPLP

COVID

(redução de *COVID-19*)

nome feminino

[Informal] [Medicina] Doença infecciosa respiratória, causada pelo coronavírus SARS-CoV-2, cujos sintomas podem incluir febre, tosse, dificuldades respiratórias e cansaço, e que, em alguns casos, pode progredir para pneumonia ou falha respiratória. = COVID-19

Nota: Também se escreve com minúsculas (*covid*). É também usado como substantivo masculino.

COVID-19

(acrónimo do inglês *coronavirus disease 2019*, doença de coronavírus 2019 [ano em que a doença foi identificada pela primeira vez])

nome feminino

[Medicina] Doença infecciosa respiratória, causada pelo coronavírus SARS-CoV-2, cujos sintomas podem incluir febre, tosse, dificuldades respiratórias e cansaço, e que, em alguns casos, pode progredir para pneumonia ou falha respiratória.

Nota: Também se escreve com minúsculas (*covid-19*). É também usado como substantivo masculino.

DLP

COVID-19, covid-19

kovid dəzɐˈnɔv(ə)

nome feminino

MEDICINA doença respiratória causada por um coronavírus (SARS-CoV-2), apresenta sintomatologia variável, desde casos assintomáticos ou formas de intensidade ligeira (cujos sintomas podem incluir febre, tosse, fadiga ou dores musculares) até situações graves (sobretudo em idosos ou pessoas com problemas de saúde preexistentes), que podem evoluir para cenários de pneumonia, falência de múltiplos órgãos e eventual morte; (inicialmente identificada na China, em 2019, atingiu estatuto de pandemia em 2020)

Do inglês *Corona virus disease 2019*, «doença de coronavírus 2019», (ano em que foi identificado o primeiro surto da doença)

Figure 5: Lexicographic articles regarding *COVID-19* (DPLP, DLP).

Looking at Figure 5, we see that the unit is classified as a feminine noun in both dictionaries, although DPLP also notes the possibility of it being used as a masculine noun: "É também usado como substantivo masculino." ["It is also used as a masculine noun."]. Although the feminine gender is recommended, our analysis of

the corpus actually attests to the fluctuation in grammatical gender. Instead of resorting to the notes field, which often includes very general information, a better structuring of the data should make the information about gender fluctuation appear in the grammatical information field itself. That is, where *nome feminino* [feminine noun] appears, lexicographers should add *nome masculino* [masculine noun], if the intention is to attest to actual usage in corpora, or simply move the note closer to the gender field, since it is purely grammatical information. It should be noted that in the DPLP case, we have two consecutive notes of different nature: one, focusing on the spelling, indicating that the form is also written in lowercase, and another with a grammatical scope, referring to the word's gender.

We now turn to the matter of the domain labels. While DLP classifies *COVID-19* as belonging to medicine, being preceded by the definition, DPLP shows two usage labels: a diaphasic label, *Informal*, and a diatechnical label, *Medicina*. As lexicographers and terminologists, we may assume the editors' intentions by including these two labels of a different nature, even though they may seem contradictory. However, we question whether an ordinary dictionary user is able to understand these labels. In our opinion, the *Medicina* label specifies the specialised domain, in which *COVID* is a term since it denotes a disease. On the other hand, the *Informal* diaphasic label in the *covid* entry may be justified by its distancing, both semantically and formally, from the original concept of < COVID-19 >, moving from its original context in the medical domain to a less specialised and more general context. However, we may question the use of this diaphasic label, since the fact that a given term becomes popularised does not necessarily mean that it starts being used in informal contexts. In the scale usually established between the informal and formal registers, the use of the reduced form *covid* can be situated in a neutral language register, or can even be used in specialised contexts. In any case, we do not see any advantage in the combination of these two labels, which may even confuse the user or raise further doubts.

Lastly, regarding term formation, DPLP highlights the fact that it is an English acronym.

Concluding our analysis, we are now in a position to answer the above-mentioned research questions:

1) DPLP and DLP record neologisms detected in Press*CoronaCorpus* and selected for analysis. It is important, however, to mention that with regard to the suffixal derivation process, at first, we had chosen as possible candidates for analysis *covidência*, *covidade* and *covidização*, due to their strong neological character. However, at the time of writing this paper, none of these lexical units was recorded in the dictionaries, so we had to exclude them.
2) None of the dictionaries have any identifying marks for neologisms or new entries.
3) None of the dictionaries shows *covid* as an element for forming new words, referring instead to the original acronym (*COVID-19*).

8 Template proposal for a lexicographic article

Following a thorough analysis of the extracted terms selected from our corpus along with the lexicographic treatment observed in the two dictionaries under study, we propose a template for the term *covid*. This proposal should bear in mind the following points:

1) the term *covid* should display its original form, both as an acronym and as a noun in lower case;
2) collocations should be included and highlighted;
3) the term *covid* should be analysed as a formative element of new words.

In Figure 6, the entry referring to the disease is presented. In the 'entry' field, the two forms identified in the corpus are registered. The lowercase form has more occurrences (47% as shown in Graph 1), followed by the acronym form. The 'part of speech' field accounts for the category (n. = noun) and the gender (f. = feminine). Gender fluctuation is not considered here, given that the use of the masculine form is to be avoided and its inclusion may confuse the end-user. This information and other related questions about orthographic rules, for example, could be given by links pointing to other lexical resources, such as spelling manuals or orthographic vocabularies, which can help clarify user questions.

The selected domain label points towards the Medical and Health Sciences domain (Costa et al. 2020). Since the lexicographic definition, starting with 'infectious disease', seems to provide sufficient clarification, the information pertaining to the domain label may be hidden from the user but still is useful to retrieve information for lexicographic purposes. In any case, its insertion is justified, as it allows the lexicographer to better control the terminology and future semantic associations made between this term and other related terms. After the lexicographic definition, which should be as objective and simple as possible, there are usage examples extracted from our corpus. In addition to these usage examples, the observed collocations are registered and also illustrated via real usage contexts.

Should there be any general observations, exemplified in Figure 6, these can be supplied under 'note'.

covid-19, COVID-19 *n.f.*

Domain label Ciências Médicas

doença infeciosa causada por um coronavírus (SARS-CoV-2)

Example: *Devido ao surto de **covid-19** vindo da variante sulafricana, a Bélgica, país onde a maioria das corridas irá para a frente, pôs em aberto o cancelamento*

Collocation **vencer a covid-19; contrair COVID-19; tratar a COVID-19**

Example: *à medida que começamos a ver o verdadeiro impacto dos tratamentos adiados em 2020 e os efeitos de longo prazo sobre aqueles que contraíram covid-19".*

Note 1. Acrónimo inglês 'coronavirus disease 2019' , 'doença do coronavírus 2019'.

Figure 6: Lexicographic article regarding *covid-19* and *COVID-19*.

Lastly, since *covid* is used as an element which forms new words, its entry as a formative is also presented in Figure 7. Following the entry, which ends in '-' in this case, clearly showing that this is a word formation element, there is respective grammatical information, as well as the domain label and related sense. The examples are depicted after that. Again, if other types of information are needed, the 'note' field can be used.

covid- *elemento de formação*

Domain label Ciências Médicas

que exprime a noção de covid (doença)

Example: *covidário; covidente; covidiano*

Note 1. Da forma reduzida covid-19, pelo acrónimo inglês coronavirus disease 2019 , 'doença do coronavírus 2019'.

Figure 7: Lexicographic article regarding *covid*.

With these two examples, we believe to have shown that a more rigorous and better segmented structuring process of lexicographic data can bring clarity to the entries and, in turn, to end users. The spelling variants (the lowercase noun and the uppercase acronym) are displayed as lemmas from the start, thereby preventing information of identical scope to be dispersed. The option to resort to the domain label does not necessarily mean that this unit is only used in specialised contexts. Instead, it helps to frame the unit within a previously outlined domain taxonomy.

Although corpus-based examples, and mainly collocations, play a key role in helping the user observe those units in real usage contexts, this seems not to be particularly valued by both DPLP and DLP. Notes, too, will always be helpful in providing other types of information which may be useful to the end user. Ultimately, the innovative contribution of our approach is to introduce an entry for *covid* as a formative element.

9 Concluding notes

In this paper, we analysed new lexical units which have arisen in European Portuguese amidst this pandemic situation. These new lexical units – neoterms and neologisms – emerged for two main reasons. Firstly, the experts needed to designate new concepts which appeared within a specialised context. Secondly, there is also the need to transfer information produced by experts to a non-expert audience. This knowledge transfer is carried out not only by experts, especially from public health settings, but also by journalists and other authors, via discourse production. When this knowledge is transferred to a non-expert audience, information is lost, given that the latter does not have the required knowledge to understand the specialised content of that message. On the other hand, journalists, also a non-expert group, interpret

specialised texts and try to reproduce the information, and therefore information is very likely to get lost, in what concerns both rigour and precision.

While terminology use is present throughout these communication scenarios, neoterms sometimes lose their status and become simple neologisms, thereby leading to determinologisation processes, since there is a context shift which entails a loss of their specialised nature. The extremely fast pace at which new units emerge, either neologisms or terms, has a strong impact on the lexicographer's work. The urgency of publishing neologisms of high daily frequency in real time, as well as the need to meet the research requirements of general language dictionary users, often lead to a certain hastiness, not giving lexicographers the opportunity to conduct a thorough analysis and subsequent validation of their data.

On the other hand, dictionaries can be 'descriptive' or 'prescriptive/normative', establishing the model to follow. Prescriptivism is an approach that attempts to determine the rules of correct usage of a language, while descriptivism is an approach that analyses and describes how the speakers of a language actually use it. Concerning the dictionary as a language model, descriptive guidance has become more usual, a process facilitated by the fact that lexicographers can access increasing amounts of corpora to support their descriptions. We maintain this approach, even though we consider that descriptive dictionaries benefit from a certain normative tone (hence we do not consider the occurrences of *covid* as a masculine noun in our proposal). Users ultimately resort to dictionaries to clarify their doubts and to ensure a correct usage of language.

As stated by Nová (2018: 397), "there is probably no universal way to treat determinologized words, but many of them need a special approach". Some fields can be used for this purpose, as is the case of the notes field, as we have shown. However, it is necessary to take into account that we are dealing with general language dictionaries, i.e., the notes should never be too long.

Aware of the difficulty of registering neologisms in general language dictionaries, especially in the context of an on-going pandemic in which hundreds of specialised units are subject to daily analysis, we found that Portuguese lexicographic resources would greatly benefit from presenting information extracted from analysis corpora and gradually accounting for observed phenomena, such as the determinologisation of terms. In this sense, our paper intends to be a contribution to the advancement of national lexicography.

Bibliography

Bauer, Laurie (2003): *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.

Costa, Rute (2017): Les collocations terminologiques. In: *Provas de agregação, Lexicologia, Lexicografia, Terminologia*. Lisbon: FCSH UNL.

Costa, Rute/Carvalho, Sara/Salgado, Ana/Simões, Alberto/Tasovac, Toma (2020): Ontologie des labels de domaines appliquée aux dictionnaires de langue générale. In: Blanco, Xavier (ed.): *La lexicographie en tant que méthodologie de recherche en linguistique. Langue(s) et Parole – Revue de Philologie Française et Romane* (5).
DLP = *Dicionário da Língua Portuguesa* [online]. Infopédia, Porto Editora, 2003–2021: http://www.infopedia.pt/dicionarios/lingua-portuguesa/ (last access: October 5, 2021).
DPLP = *Dicionário Priberam da Língua Portuguesa* [online]. Priberam, 2008–2021: http://www.priberam.pt/dlpo/ (last access: October 5, 2021).
Galisson, Robert (1978): *Recherches de lexicologie descriptive, la banalisation lexicale: le vocabulaire du football dans la presse sportive, contribution aux recherches sur les langues techniques*. Paris: F. Nathan.
Guilbert, Louis (1975): *La créativité lexicale*. Paris: Librairie Larousse.
ISO 1087: 2019: *Terminology Work and Terminology Science – Vocabulary*. Geneva: ISO.
Lino, Maria Teresa (2019): Neologia e neonímia em língua portuguesa. In: *Linha D'Água* (online), v. 32, n. 3, p. 9–23. Brazil: São Paulo.
Lombard, Alizée/Huyghe, Richard/Gygax, Pascal (2021): Neological intuition in French: A study of formal novelty and lexical regularity as predictors. In: *Lingua*, Volume 254.
Marquilhas, Rita (2014): Fenómenos de mudança na história do português. In: Raposo, Eduardo/Nascimento, Maria Fernanda/Mota, Maria Antónia/Seguro, Luísa/Mendes, Amália (eds.): *Gramática do Português*. Lisboa: Fundação Calouste Gulbenkian, 17–48.
Meyer, Ingrid/Mackintosh, Kristen (2000): When terms move into our everyday lives. An overview of de-terminologization. In: *Terminology*, Volume 6–1, Amsterdam/Philadelphia: John Benjamins, 111–138.
Nová, Jana (2018): Terms Embraced by the General Public: How to Cope with Determinologization in the Dictionary? In: *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana, Slovenia, 17–21 July 2018, 387–398.
Pruvost, Jean/Sablayrolles, Jean-François (2003): *Les néologismes*. Paris: Presses Universitaires de France.
Rio-Torto, Graça/Rodrigues, Alexandra Soares/Pereira, Isabel/Pereira, Rui/Ribeiro, Sílvia (2013): *Gramática Derivacional do Português*. Coimbra: Imprensa da Universidade de Coimbra. ISBN 978-986-26-0640-8.
Rondeau, Guy (1984): *Introduction à la terminologie*. Chicoutimi: Gaëtan Morin éditeur.
Sinclair, John (1996): EAGLES. Preliminary recommendations on Corpus Typology. EAG–TCWG–CTYP/P. http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html

Mireille Vale, Rachel McKee

Neologisms in New Zealand Sign Language: A case study of COVID-19 pandemic-related signs

1 NZSL background and lexicography

New Zealand Sign Language (NZSL)[1] is estimated to be used by 3,000–5,000 Deaf people in New Zealand, with a larger group of just over 20,000 New Zealanders able to “have a conversation about a lot of everyday things” in the language.[2] Prior to the development of interpreting services in the 1980s and acceptance of NZSL in education from the 1990s, NZSL was used mainly for communication in private, social domains, which restricted the size and fields of lexicon.

Linguistic documentation of NZSL began in the mid-1980s (Collins-Ahlgren 1989, Levitt 1986). Early lexicographic efforts culminated in the print *Dictionary of New Zealand Sign Language* (Kennedy et al. 1997) followed by a *Concise Dictionary of New Zealand Sign Language* (Kennedy et al. 2002) comprising the 2,000 most frequent signs. These print dictionaries were amongst the first corpus-based signed language dictionaries that used data from signed language as the source of lexicon rather than being a translational glossary from the spoken language. Lexical documentation was based on the systematic analysis of video recorded, mainly spontaneous discourse around elicited / guided topics. An extensive community validation process was undertaken before signs (including variants) were entered in the dictionary.

The existence of these dictionaries contributed to legal recognition of New Zealand Sign Language in 2006 (McKee 2006). Official language status and disability access measures have subsequently made NZSL more visible in public domains. Deaf NZSL users increasingly participate in wider social, political, occupational and educational domains, leading to rapid lexical development of NZSL in these fields. This parallels lexical expansion seen in the national indigenous language, Te Reo

1 It is conventional in linguistics literature to use the phrase ‘signed languages’ when referring to languages in this modality in a general or collective sense (cf. ‘spoken languages’ or ‘written languages’). However, the proper names of specific national languages in English take the form ‘(New Zealand / American / British . . .) Sign Language’.

2 http://www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-culture-identity/languages.aspx (last access: 10 June 2022).

Mireille Vale, Victoria University of Wellington, Deaf Studies Research Unit, PO Box 600, Wellington 6140, New Zealand, e-mail: micky.vale@vuw.ac.nz
Rachel McKee, Victoria University of Wellington, Deaf Studies Research Unit, PO Box 600, Wellington 6140, New Zealand. e-mail: rachel.mckee@vuw.ac.nz

https://doi.org/10.1515/9783110798081-013

Māori, as an outcome of recognition and revitalisation (Harlow 1993). The Deaf community's participation in new domains is typically mediated by interpreters, who are challenged by lexical inequivalence between English and NZSL.

Representing a visual-gestural language with static images (in the absence of a written form) is a key challenge in signed language lexicography (McKee/McKee 2013). Improvements in digital media and data storage enabled the creation of the *Online Dictionary of New Zealand Sign Language* (ODNZSL) with video content. Taking the dictionary online included revising and revalidating existing data and adding further entries and video material (with corpus-derived but edited example sentences). Further entries have been added in batches, with the most recent update in 2017. By signed language dictionary standards, the 6,000 or so entries in the ODNZSL make it a reasonably large and comprehensive dictionary; signed language lexicons are relatively small due to limited lexicalisation, the capacity of productive forms to express novel context-dependent meanings, and the fact that signed languages were historically used in limited domains (Johnston 2012). The dictionary is a general-purpose dictionary primarily aimed at L2 learners rather than at the Deaf community, and for this reason the initial focus was on documenting high frequency signs. A user study found that use of dictionary content in teaching materials is a primary function for Deaf NZSL users and that it may also have an authoritative / standardising role, but is rarely used by Deaf NZSL users to look up the meaning or form of unknown signs (Vale 2015). Corpus work in NZSL has been undertaken in projects from the 1990s, but annotation of signed language corpora is complex and labour intensive, and the dictionary does not have access to a highly contemporary corpus from which to source current neologisms.

To leverage the Deaf community's increasing online presence, the web-based platform NZSL Share was launched in March 2020 to crowdsource new and previously undocumented signs, and to encourage community validation of these signs. The platform allows users to upload sign videos, comment on videos and agree or disagree with (often new) signs being proposed. It is managed by the research team that maintains the ODNZSL, which includes the authors. NZSL Share is being used by individuals as well as Deaf community groups to record and share signs of a specialist nature (e.g., school curriculum signs). NZSL Share now has close to 50 actively contributing members. Its launch coincided with the 2020 COVID-19 outbreak in New Zealand and so some of the first signs contributed were COVID-19-related, which are the focus of this paper.

2 COVID-19 in New Zealand

The first COVID-19 case in New Zealand was reported on 28th February 2020 and by the end of March, the entire country was required to comply with a full lockdown

(known locally as Alert Level 4) with the aim of eliminating COVID-19 from the community. During this time, the government and public health officials broadcast daily updates through television, radio and print media. It was vital that these communications reached all communities rapidly and so NZSL translation of print information was commissioned through agencies associated with the Deaf community, and NZSL interpreters were deployed at official televised briefings (also posted online). Interpreters and translators were thus at the front line of communicating new information to the NZSL community, always working under time pressure, with few reference sources and, along with the rest of the population, encountering new information and jargon as the pandemic unfolded daily. As such, interpreters and translators become de facto language innovators – generating translations and establishing terms ahead of Deaf community usage. Translation-driven lexical innovation is common when a minority language is used to translate information in public domains, as with Irish for example (Ní Ghearáin 2011). The Deaf NZSL community could not contribute greatly to creation of terminology at the outset of the COVID-19 pandemic because they were also grappling with the new information and concepts conveyed to them via translation. Furthermore the whole population was isolating at home which restricted discourse in NZSL at a community level about COVID-19, beyond online video interaction. While novel lexicon is the focus of this paper, terminology was just one of many significant challenges in mediating information to the NZSL community during the pandemic.

3 Method

We aimed to investigate translators' and interpreters' strategies for dealing with the demands of new terminology and lexical inequivalence, and their observations about the conventionalisation and dissemination of COVID-19-related signs that they used. We also wanted to explore how and when such neologisms could be entered in the ODNZSL. To gather data, we catalogued signs related to COVID-19 that were contributed to NZSL Share, and conducted two focus group interviews with: (1) interpreters who had interpreted briefings on TV (hearing L2 NZSL users, professionally trained), and (2) translators who had produced NZSL versions of public health information bulletins (Deaf L1 NZSL users, bilinguals with experience but no formal training).

Focus group interviews sought to elicit vocabulary that prompted innovation in translation, the strategies that participants used to convey new terms and concepts on the spot, and observations around the development and dissemination of coronavirus-related signs among interpreters/translators and into the wider community.

4 Findings

4.1 Novel terminology in COVID-19 related information

From the signs contributed to NZSL Share and from interview data, we identified types of novel terms and phrases in English that created challenges for translators and interpreters, and which therefore could trigger neologisms in NZSL. These included not only COVID-19-related terms but also adjacent vocabulary relating to economic and social aspects of the pandemic response. We loosely categorise this vocabulary below in terms of reasons for lexical challenges in NZSL (see Table 1).

Table 1: Categories of challenging terms and phrases in translation.

A. Medical / testing related terms – English (technical)
Antibody, community transmission, coronavirus, covid-19, covid-positive, covid-negative, dose, epidemic, epidemiological link, genome testing, herd immunity, nasopharyngeal swab, negative pressure room, pandemic, PPE, screening, strain, vaccine, vaccine rollout, virus
B. Other new / extended / reframed concepts – (NZ) English
alert levels 1–4, bubble, case, casual (+) contact, close contact, eliminate, eradicate, essential services, lockdown, mask, MIQ/ managed isolation, places of interest, quarantine, self-isolation, social distancing, team of five million, trans-Tasman bubble
C. Lexical gaps in NZSL / difficult to translate concepts
border, closing the border, hygiene, fiscal, mortgage holiday, notice (official Government notice), Reserve Bank, rent freeze, road block, support package, symptoms

Firstly, many terms that frequently occured in the government information and media briefings were **technical medical terms** already in use in English (with the exception of COVID-19). Some of the terms in Category A might be reasonably common (e.g., *epidemic*, *vaccine*, *immunity*) but others would previously have had limited use beyond the medical /scientific community (e.g., *genome sequencing*, *negative pressure room*).

Category B consists of terms that were either neologisms in NZ English or that were used in an **extended or specific sense in relation to COVID-19** (such as *lockdown*, *alert levels*, *bubble*, *essential workers*).

Finally, Category C contains terms in the source language that were not new or not directly related to COVID-19, but were nevertheless challenging because no equivalent signs exist in NZSL (such as *symptoms*, *border*). Some of these **lexical gaps** arose in relation to jargon around economic and social policy responses (e.g., *Reserve Bank*, *employment support package*, *mortgage freeze*).

4.2 Strategies for lexical innovation and types of resulting NZSL coinages

Known strategies for lexical innovation include semantic extension; coinage of new words through language-internal mechanisms such as derivation or compounding; and drawing on language-external resources, as calques or direct loans. The extent to which specific strategies are used and are deemed acceptable may vary according to the preferences of the language community (Jernudd 2013).

Proposed COVID-19 related signs contributed to NZSL Share as well as translational equivalents discussed by interpreters and translators in our focus groups include examples of both language-internal and language-external lexical innovation strategies (see Table 2). These examples reflect processes of sign creation found in the NZSL lexicon generally, as evident from signs entered in the ODNZSL and from contributions to NZSL Share.

Table 2: Types of lexical innovation in NZSL coinages and translational equivalents.

Lexical innovation strategy	NZSL equivalent
NZSL Linguistic resources	
Paraphrasing (circumlocution)	– MIQ (managed isolation and quarantine > STAY^HOTEL[3] – pandemic > ILL^SPREAD^WORLD – eliminate > COVID^STOP
Hypernyms expanded to a list	– symptoms > FEVER, SORE-THROAT, COUGH – PPE > MASK, GLOVES, APRON – Hygiene practices > WASH-HANDS, COUGH-INTO-ELBOW
Grammatical restructuring, e.g., nominal referents > verb phrases	– Trans-Tasman bubble > AUSTRALIA NZ PLANES-FLY-reciprocal – transmission > PERSON^PASS-ON-multiple
Productive morphology (depicting/ visually motivated)	SOCIAL-DISTANCING, QUARANTINE, ALERT-LEVELS
Semantic extension	– SPREAD > ‘EPIDEMIC’ – GROUP > ‘CLUSTER’
External linguistic resources	
Calques (from English)	ESSENTIAL^OCCUPATIONS, LOCKDOWN, COVID^POSITIVE, CLOSE^CONTACT
Loan sign (from other sign languages)	CORONAVIRUS

3 In this paper we follow the convention of representing lexical signs with capitalised English glosses.

We note that a large proportion of 'signs' entered in NZSL Share are actually phrasal (multi-sign) translations of a concept, rather than lexicalised coinages. A further common strategy to express novel meaning in signed languages is the use of productive morphology to construct complex predicates, often motivated by visual properties of the referent. For example, Figure 1 shows common productive constructions in which the upright index finger represents a person, and Figure 2 shows how the same productive elements are used in the coinage of an equivalent for *social distancing*.



Figure 1: Complex predicates using the productive PERSON handshape.

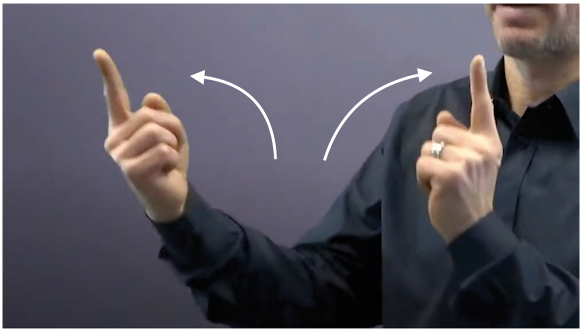


Figure 2: 'SOCIAL-DISTANCING'.

The use of such strategies in relation to COVID-related lexical innovation is consistent with an investigation of health-related terminology in Auslan (Australian Sign Language), in which relatively few terms were found to have a conventional lexicalised form, but rather were expressed by depicting strategies (Major et al. 2012).

Polysemy is prevalent in NZSL, and accordingly, lexical extension is used liberally for expressing new COVID-19 related meanings – by attaching a novel contextual meaning to an existing sign by mouthing the corresponding English term with the sign (McKee 2007).

Although unrelated in both modality and structure to the dominant spoken languages that surround them, signed languages are subject to constant influences arising from close language contact. Calques from the spoken language are therefore

relatively common, especially for two-part terms or phrases, as reflected in existing NZSL dictionary entries such as *open-minded*.[4]

Contact and borrowing between national signed languages is a common phenomenon. The visual-gestural production modality of signed languages means that they tend to share more phonological and morphological material (especially visually motivated elements) than spoken languages, which facilitates the sharing of lexicon across language boundaries (Quinto-Pozos/Adam 2015). Borrowing in the context of COVID-19 is therefore consistent with a general trend for NZSL users to readily adopt vocabulary from other signed languages to fill lexical gaps or expand the lexicon, and online exposure to texts in other signed languages seems to be accelerating this trend (McKee/McKee 2020). In the current study we identified four loans from overseas signed languages, which were apparently acquired from foreign online sources. Chief among these is the sign CORONAVIRUS / COVID, which is anecdotally said to have originated in Japan and was widely adopted into many signed languages early in the pandemic.

4.3 Interpreters and translators' use of lexical innovation strategies

Interpreters and translators may be agents of language change by introducing and disseminating neologisms to the target language community through their renditions (Lenihan 2018). The same typical lexical innovation resources discussed above are available as translation strategies in response to novel concepts or source text neologisms, or introduced into the target text as idiosyncratic usage by the interpreter/translator (Niska 1998). Which strategies are prevalent in translations is affected by the general trends of the target language, but may also vary according to individual interpreters (Van Obberghen 2016). The constraints of simultaneous interpreting (or short-notice translation) may also influence the use of certain strategies. For example, calques from English may be a default (but temporary) response when first hearing a neologism or unfamiliar term.

Interpreters and translators in our focus group interviews demonstrated a high level of awareness and concern about their potential influence on NZSL language change. Although as mentioned above, some new coinages may be the direct result of the demands of working under time pressure, our research participants indicated that for the most part, they made conscious choices about the strategies they used. Primarily, they tried to avoid coining neologisms. This was largely due to the imperative to make information accessible to the Deaf community in language that would be readily understood, at a time when the Deaf community was still unfamiliar with the English term or concept and thus had no referent for new signs. For the same

4 https://www.nzsl.nz/signs/5661 (last access: 10 June 2022).

reason, our research participants were wary about using calques from English such as *COVID-positive*. Thus, the demand for lexical and translational innovation driven by novelty in the source message was in tension with considerations of comprehensibility for the target language audience – among whom health literacy is also lower than in the general population (Witko et al. 2017). Interpreters and translators reported that rather than creating new 'terms', their focus especially in early communications was to paraphrase and expand new terms with examples to maximise transparency and understanding. For example, describing someone as 'having COVID' was considered preferable to using a calque such as *COVID-positive*, because 'positive' in NZSL is more likely to be understood in its usual sense of a 'desirable attribute/attitude' rather than the intended technical sense of being present.

A further reason for avoiding neologisms was that the conditions of lockdown and time pressure to render information meant that interpreters and translators had limited access to Deaf community feedback with regard to their understanding and uptake of any such neologisms. Our research participants also reported working mainly in isolation with limited opportunities to discuss new terms in the source text with colleagues, especially at the beginning of the pandemic. As a result, translational equivalents were variable and at times idiosyncratic, causing further concern that the Deaf target audience would not be able to associate these variable translations with the new concepts and English terms. In adddition, hearing interpreters especially were conscious of language authenticity considerations as second language users of NZSL (and indeed they reported some negative comments from Deaf NZSL users in social media about their vocabulary choices or apparent innovations on the basis that they were used by non-deaf interpreters).

Together, these concerns for comprehensibility and language authenticity may have predisposed our research participants to create translational equivalents using language-internal strategies, including semantic extension, paraphrasing, grammatical restructuring (changing nominal referents to verb phrases; rendering hypernyms as a list), and using productive morphology to create 'nonce' constructions with contextual reference. Since similar signed language interpreting activity was occurring in many countries, these somewhat parallel online texts also offered a resource for browsing lexicon and translational strategies, in a few cases leading to the introduction of loan signs.

5 Discussion

5.1 Status of COVID-19-related lexical innovation

Although interpreters and translators had to exercise creativity to render a proliferation of COVID-19 related terms and concepts, many of the strategies they employed did not lead to lexical neologisms in NZSL. While extended paraphrases were

progressively shortened, and some productive forms and lists of hyponyms over time became conventionalised translational equivalents, their status as fixed lexical signs or sign phrases is uncertain. This is partly a reflection of the nature of NZSL lexical innovation processes in general. As we noted in section 4.2, productive depicting constructions convey context-specific meanings; however their reference is not fully specified when decontextualised.

Examples of productive depicting constructions used in the context of COVID-19 terms are:

- *quarantine* (indicating a fenced-off area);
- *mask* (showing a mask stretching over the nose and mouth);
- *nasopharyngeal swab* (showing a swab being inserted into the nose);
- *social distancing* (the upright index fingers of both hands depicting persons, moving apart);
- *trans-Tasman bubble* – i.e. quarantine-free travel between Australia and New Zealand (the two hands representing planes flying in reciprocal directions).

Many of these constructions can have a range of contextual meanings. For example, any fenced-off area could be described with the same construction that is used for *quarantine*, and the construction used to describe *social distancing* could also be used in the general sense of people 'standing apart' or 'avoiding' each other. The specific intended meanings of such constructions in relation to COVID-19 may not be retrievable outside of the context of the immediate translation or interpretation. Thus, it would be difficult to justify listing the form 'two planes flying in reciprocal directions' in a dictionary with the sense *trans-Tasman bubble*, for example.

Similarly, the strategy of rendering hypernyms as lists of category members may be context-dependent and even when largely conventionalised, such lists cannot be said to have fixed lexical status (Kennedy et al. 1997).

Some terms had a lexical character, but had variable form across different individuals and contexts of use. An example is *border*, which was hitherto a low frequency concept in NZSL discourse, perhaps in the absence of land borders in New Zealand. (Interestingly, the sign that appears in the ODNZSL[5] is exemplified by a sentence about the border between USA and Mexico, suggesting that this sign is seldom used with local reference.) Interpreters explained that their translations of *border* and *border workers* in the COVID briefing situation varied according to the specific referent – e.g., sea port, airport, or state line (in reference to travel restrictions within Australia). When a generic term was unavoidable (e.g., a phrase such as *border closure*), they indicated a line or boundary in various ways, but doubted that these varying forms would become conventionalised given low frequency use beyond this situation.

5 https://www.nzsl.nz/signs/480 (last access: 10 June 2022).

5.2 Implications for NZSL lexicography

The lexicographical treatment of NZSL neologisms, including new coinages arising from the COVID-19 context, has to be considered against the background of our past and present lexicographical practices and the purpose and format of the ODNZSL. This dictionary and its precursors, as mentioned previously, used corpus evidence and community validation processes in the documentation of high frequency signs for general purposes. It is clear from the findings of the current study as well as from our ongoing lexicographical work that although similar language innovation processes are at work, many recent neologisms are of a different nature to previously documented high frequency signs. Not only are they used and recognised by much smaller subsets of the language community, but they often arise from interpreted or translated English material (in specialised areas) rather than spontaneous community usage.

In recent years, the ODNZSL has broadened its scope and has entered a number of NZSL neologisms in specialist areas such as school science and mathematics curricula, linguistics, and local place names. Expanding an existing online dictionary with neologisms requires changes in methodology to collect and validate data, as well as extensive revisions to the web application to meet diverse user needs (Expertisecentrum Vlaamse Gebarentaal n.d.). Although broadening the scope of the ODNZSL has already required some procedural changes, it is likely that a number of core principles regarding the addition of new entries will remain unchanged. When we asked our research participants how we could determine which, if any, of the COVID-19 related translational 'innovations' should be entered in the dictionary, their suggestions were consistent with these core criteria:

- The coinage should be **a fixed lexical sign or sign phrase**, not a one-off coinage or construction that only has reference within the immediate context.
- The coinage must have **longevity and transferability** – i.e., use beyond the context of COVID-19 briefings. Since the initial 2020 lockdown in New Zealand, the focus of COVID-19 reporting and discourse has continually changed and some of the original terms used in English and NZSL have altered or reduced in frequency. While such coinages might be of historical interest, the primary functions of the ODNZSL would not be well served by the inclusion of short-lived coinages that are no longer current.
- **Wider use:** there must be evidence of the sign being used by the Deaf community, beyond just translators and interpreters.

Very few of the terms mentioned in our findings meet these criteria. Perhaps unsurprisingly, the most cited case of an established new 'sign' in our data is the loan sign CORONAVIRUS / COVID, which now shows evidence of widespread community usage in New Zealand. In addition, a small number of productive depicting

constructions (such as MASK and NASOPHARYNGEAL-SWAB) are sufficiently lexicalised to consider entering in the dictionary.

Over time, further COVID-19 related signs may stabilise and meet these criteria, while other terms may fall out of use or not be taken up by the Deaf community. We will continue to monitor the coinages discussed in this paper as part of a wider research project investigating recent vocabulary growth in NZSL and the prevalence of language-internal vs. language-external factors in new sign creation. Since it is not possible to automatically extract relevant terms from video texts, we foresee a significant role for NZSL Share as a crowdsourced repository for new terms.

Due to the circumstances in 2020 and 2021, it has not yet been possible to effectively recruit community contributors to NZSL Share. As a result, NZSL Share was of limited use as a tool or strategy for rapid sharing of neologisms during the first wave of the pandemic. The rate of new terminology quickly outstripped community capacity to innovate and record equivalents. In practice, the interpreters on daily TV briefings became the most visible daily source of new vocabulary or phraseology. Some coinages recorded in NZSL Share were found to be idiosyncratic and novel, thus were not useful to interpreters and translators to communicate to a wide community audience (e.g., an individual's coinage for *antibody*). Translators and interpreters reported that while they looked for vocabulary in NZSL Share and the ODNZSL, more frequently they referenced each other's work to standarise their vocabulary usage as far as possible. Thus, the process of vocabulary creation and dissemination became somewhat self-referential without an effective standardising or advisory mechanism, which was not possible to organise effectively under the restricted circumstances in which this process unfolded. In spite of these limitations, community reactions to NZSL Share have been very positive and uptake by individuals and groups (such as the national Deaf education provider) is gradually increasing as we continue to promote the platform.

We note that other signed language dictionaries are grappling with similar methodological and lexicographical issues with regard to new signs. The Woordenboek Vlaamse Gebarentaal (Flemish Sign Language online dictionary) now includes an interface to allow for crowdsourced contributions; an expert validation committee meets several times a year to discuss such contributions and other neologisms identified through linguistic research. The validation status of signs in various regions is marked on entries in this online dictionary, with unvalidated signs shown as 'not yet known'. This approach allows the Flemish Sign Language dictionary to make new sign terminology (including COVID-19 related signs) available online quickly.

Whilst we acknowledge the potential benefits of documenting the NZSL lexicon in one place, we anticipate that NZSL Share will be maintained as a separate website at present. As a separate platform, NZSL Share can include community contributions that do not (yet) meet the criteria of fixed lexical form and longevity as well as signs that are not typically included in a non-specialist dictionary, such as brand names or name signs used with the Deaf community to refer to public figures. It

will provide a forum for consensus-building and dissemination of new signs in the NZSL-using community, in the absence of a language planning body or expert committee. The platform will also allow us to adapt our processes to include online validation with specific groups of language users. At the same time, the ODNZSL can continue to be a trusted resource of a community-validated lexicon, and a consistent format can be maintained for dictionary entries, which include learner-focused example sentence videos as well as grammatical and user information that require an editorial role.

6 Conclusion

This case study of COVID-19 related lexical innovation in NZSL has shown that the main driver for new terminology has been live interpreting and translation of government and public health information. There has been rapid generation of new coinages in both directly COVID-19 related and adjacent fields, using both language-internal strategies (semantic extension, paraphrasing, grammatical restructuring, productive morphology) and language-external resources (calques and loans). Interpreters and translators as the primary source of this lexical innovation showed a high level of concern for language authenticity and comprehensibility, which influenced the strategies they chose to render new terms and concepts into NZSL.

Very few of the new coinages meet the criteria for being entered in the ODNZSL, due to the uncertain lexical status of some constructions, variable and at times idiosyncratic usage, and difficulties in determining dissemination and adoption of new signs in the wider NZSL community.

While COVID-19 related lexical development therefore will not have an immediate impact on the ODNZSL, this study has implications for the role(s) and format of the dictionary and highlights potential changes required in our lexicographic processes to account for the nature of NZSL neologisms.

Although it was found to be of limited immediate use as a tool for rapid sharing of neologisms during the first wave of the pandemic, it is expected that the crowdsourcing platform NZSL Share, launched in 2020, will facilitate collection, community validation and dissemination of sign neologisms.

Bibliography

Dictionaries

A Dictionary of New Zealand Sign Language. Ed. by Kennedy, Graeme et al. Auckland, New Zealand: Auckland University Press with Bridget Williams Books 1997.
A Concise Dictionary of New Zealand Sign Language. Ed. by Kennedy, Graeme D. et al. Wellington, New Zealand: Bridget Williams Books 2002. (2002).
ODNZSL = *Online Dictionary of New Zealand Sign Language*. Ed. by McKee, David et al. [https://www.nzsl.nz, last access: August 24, 2021].
NZSL Share. Maintained by Deaf Studies Research Unit. Wellington: Victoria University of Wellington (n.d.) [https://www.nzslshare.nz, last access: August 24, 2021].
Woordenboek Vlaamse Gebarentaal. Ed. by Expertisecentrum Vlaamse Gebarentaal. (n.d.). [https://www.vlaamsegebarentaal.be/methodologie#methodologie, last access: August 25, 2021].

Monographs, edited volumes and articles in edited volumes or journals

Collins-Ahlgren, Marianne (1989): *Aspects of New Zealand Sign Language*. Victoria University of Wellington.
Harlow, Ray (1993): Lexical expansion in Māori. In: *The Journal of the Polynesian Society* 102(1),99–107.
Jernudd, Björn H. (2013): Linguistic sources for terminological innovation: Policy and opinion. In: Rubin, Joan/Jernudd, Björn H./DasGupta, Jyotirindra/Fishman, Joshua A./Ferguson, Charles A. (eds.): *Language Planning Processes*. Berlin / Boston: De Gruyter Mouton, 215–236. DOI:10.1515/9783110806199.215.
Johnston, Trevor A. (2012): Lexical frequency in sign languages. In: *Journal of deaf studies and deaf education* 17(2),163–93. DOI:10.1093/deafed/enr036.
Lenihan, Aoife (2018): Language Policy and New Media: An Age of Convergence Culture. In: Tollefson, James W./Pérez-Milans, Miguel (eds.): Oxford Handbook of Language Policy and Planning. 1–21. DOI:10.1093/oxfordhb/9780190458898.013.33.
Levitt, Dan (1986): *An introduction to New Zealand Sign Language*. Auckland: National Foundation of the Deaf.
Major, George, et al. (2012): Exploring lexical gaps in Australian Sign Language for the purposes of health communication. In: *Communication and Medicine* 9(1),37–47.
McKee, Rachel (2007): Hand to mouth : The role of mouthing in NZSL. In: *Across the Board: Australian Association of Sign Language Interpreters (Victoria)* 2(4),3–12. DOI:10.13140/2.1.2032.0965.
McKee, Rachel (2006): The eyes have it! Our third offical language: New Zealand Sign Language. In: *Journal of New Zealand Studies* NS-4-5, 129–148.
McKee, Rachel/McKee, David (2020): Globalization, hybridity and vitality in the linguistic ideologies of New Zealand Sign Language users. In: *Language and Communication* 74, 164–181.
McKee, Rachel/McKee, David (2013): Making an online dictionary of New Zealand Sign Language. In: *Lexikos* 23, 500–531.

Ní Ghearáin, Helena (2011): The problematic relationship between institutionalised Irish terminology development and the Gaeltacht speech community: dynamics of acceptance and estrangement. In: *Language Policy* 10, 305–323. DOI:10.1007/s10993-011-9216-5.
Niska, Helke (1998): *Explorations in translational creativity: Strategies for interpreting neologism*. Stockholm University.
Quinto-Pozos, David/Adam, Robert (2015): Sign language contact. In: Bayley, Robert, et al. (eds.): *Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press, 1–16. DOI:10.1093/oxfordhb/9780199744084.013.0019.
Vale, Mireille (2015): A study of the users of an online sign language dictionary. In: Kosem, I., et al. (eds.): *Electronic Lexicography in the 21st Century: Linking Lexical Data in the Digital* Age. *Proceedings of the ELex 2015 Conference*, *11–13 August 2015, Herstmonceux Castle, United Kingdom*. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., 281–303.
Van Obberghen, Mathieu (2016): *Onderzoek naar het tolken van neologismen: Analyse en reflectie*. Brussel: Vrije Universiteit Brussel.
Witko, Joanne, et al. (2017): Deaf New Zealand Sign Language users' access to healthcare. In: *New Zealand Medical Journal* 130(1466),53–61. DOI:10.1161/01.STR.22.9.1117.

Franck Sajous

Using Wiktionary revision history to uncover lexical innovations related to topical events: Application to Covid-19 neologisms

1 Introduction

In April and July 2020, two extraordinary updates of the *Oxford English Dictionary* (OED) focused on the neologisms related to the Covid-19 pandemic. The responsiveness of the OED was made possible by the ability of its team to monitor, analyse and report quickly a sudden inflow of lexical changes. This ability, while not unique, is not prototypical in the lexicographic landscape. Corpus lexicography obviously requires corpora, but also tools to process and query them and sufficient person-hours. Fulfilling these standard requirements simultaneously, however, is no trivial task. The tools are not an issue as far as lexical creations are concerned. Building a headword list is indeed not considered a “hard part of lexicography” (Kilgarriff 1998) and detecting formal neologisms to update a nomenclature only requires “simple maths” (Kilgarriff 2009). Identifying semantic changes is more challenging. Clustering algorithms have been devised by Cook et al. (2013) while recent approaches use diachronic word embeddings (Fišer/Ljubešić 2018). These methods enable the detection of cultural shifts and linguistic drifts (Hamilton et al. 2016) but error rates are generally high. Another issue is that prediction-based models are appropriate for the detection of semantic changes over long time spans (decades or centuries) in very large corpora but they rarely perform well with shorter time units and smaller corpora (Kutuzov et al. 2018). On the corpus side, appropriate text collections to be used as input for the tools (i.e. diachronic corpora updated on a regular basis) are – sadly – not publicly available for most languages. Lastly, corpus lexicography also requires substantial manpower – ideally, trained lexicographers – to analyse vast amounts of data in a reasonable timeframe. Most institutions however, whether private or public, rarely have the manpower and the time they would like. The limitations are bound to the conditions of dictionary production rather than being intrinsic to corpus-based or corpus-driven approaches, as Landau (2001: 323) explained:

Acknowledgements: My thanks go to Basilio Calderone for checking the statistical analyses. The parsing of Wiktionary revision logs was performed using the OSIRIM platform, which is administered by IRIT and supported by CNRS, the Region Midi-Pyrénées, the French Government and ERDF.

Franck Sajous, CLLE – CNRS & Université de Toulouse 2, Maison de la Recherche – 5, allées Antonio Machado – F – 31058 Toulouse Cedex 9, e-mail: franck.sajous@univ-tlse2.fr

https://doi.org/10.1515/9783110798081-014

> Dictionaries are not written in a vacuum, but by people working under the pressure of time. It sometimes seems to me that as technology has improved the speed and power with which we can examine the language, the pressures to produce quickly and with fewer staff have kept pace, so that on balance nothing is accomplished any faster or better. The expectations of management seem to rise at the same rate as the speed and power of the computer increase [. . .] Corpora can be used well or they can be used badly. Time pressures too often push the lexicographer to cut corners to avoid time-consuming analyses. It really doesn't do much good having a good corpus with marvelous analytical tools if they aren't used.

Time pressure and manpower are conversely not an issue in collaborative projects such as Wiktionary, which relies on massive online contributions performed by crowds of amateurs, not on corpus-driven analysis. Despite this questionable approach to lexicography and the resulting weaknesses described, inter alia, by Hanks (2012) and Rundell (2017), the exhaustiveness and the responsiveness of Wiktionary can be leveraged to detect lexical changes. Sajous et al. (2018) showed how swiftly the crowds are likely to detect formal and semantic neology. For example, in 2017, 73% of the entries added to the OED were already recorded in Wiktionary, whose median lead time was 4 years.

In the present contribution, I investigate if and how the English and French editions of the Wiktionary collaborative dictionary can be used as a corpus for real time neology watch. This option is envisaged as a stopgap, when no satisfactory corpus is available. Wiktionary can also prove useful in addition to standard corpus analysis, to minimize the risk of overlooking new coinages and new senses. Since the collaborative dictionary's quest for exhaustiveness makes the manual inspection of the new additions unreasonable (more than 31,000 English lemmas and 11,000 French lemmas entered the nomenclature in 2020), identifying the possibly relevant headwords is an issue. The solution proposed here is to use Wiktionary revision history to detect the (new or existing) entries that received the greatest number of modifications. The underlying hypothesis is that the most heavily edited pages can help identify the vocabulary related to "hot topics", assuming that, in 2020, the pandemic-related vocabulary ranks high. I used two measures introduced by Lih (2004), whose aim was to estimate the quality of Wikipedia articles: the so-called *rigour* (number of edits per page) and *diversity* (number of unique contributors per page). In the present study, I propose to adapt the rigour and diversity metrics to Wiktionary in order to identify the pages that generated a particular stir, rather than to estimate the quality of the articles. I do not subscribe to the idea that – in Wiktionary – more revisions necessarily produce quality articles (more revisions often produce *complete* articles). I therefore adopt Lih's notion of diversity to refer to the number of distinct contributors, but leave out the name *rigour* when it comes to the number of revisions. Wolfer and Müller-Spitzer (2016) used the two metrics to describe the dynamics of the German and English editions of Wiktionary. One of their findings was that the number of edits per page is correlated with corpus word frequencies. The variation in number of page edits should therefore reflect to some extent the variation of corpus word

frequencies. Renouf (2013) established a relationship between the fluctuation of word frequencies in a diachronic corpus and various neological processes. In particular, she illustrated how specific events generate sudden frequency spikes for words previously unseen in the corpus. For instance, *Eyjafjallajökull*, the – existing – name of an Icelandic glacier, appeared in the corpus when the underlying volcano erupted in 2010 and disrupted air traffic in Europe. In order to check if the same phenomenon occurs when using Wiktionary edits instead of corpus frequencies, I manually annotated the most frequently revised entries (according to various ranking scores) with the binary tag: "related to Covid-19" (yes/no). The annotations were then used to test the ability of various configurations to detect relevant headwords from the English and French Wiktionary, namely Covid-19 neologisms and related existing words that deserve updates.

2 Methodology

Scrutinising Wiktionary offers several opportunities for collecting Covid-related neologisms quite easily, depending on the language edition, and one's ability to automatically process the content of the dictionary. First of all, the *Coronavirus* category[1] of the English Wiktionary included 52 headwords on January 1st, 2021 and 124 in June. The English Wiktionary also has a category named *Hot words newer than a year*.[2] These words are described in Wiktionary as "presumably failing the criteria for inclusion on the *spanning less than a year* requirement", but are kept, according to Wiktionary, "because they have become widely used in that short time". Which is precisely the subject of the present study. In January 2021, the category included 94 words, 26 of which were not English. 79% of the English words (54 out of 68) were related to Covid-19. This observation is encouraging in that it suggests that the 2020 hot words are those related to the pandemic. Relying on the two categories mentioned is probably a good start, but by no means a satisfactory solution. First, some headwords that would deserve to be classified in these categories are not. Second, headwords that are related, but not specifically, to Covid-19, do not necessarily fit into these categories. Third, the goal of the present study was to develop a method for discovering topical neologisms that can be adapted to other language editions and other topics. In the French Wiktionary, there is no such thing as a "coronavirus" or a "hot words" category, and such categories will not make it possible to discover neologisms related to other "hot topics" in the future. Looking for some patterns (*covid, corona*, etc.) in the headword list, the definitions and the usage

1 https://en.wiktionary.org/wiki/Category:en:Coronavirus
2 https://en.wiktionary.org/wiki/Category:Hot_words_newer_than_a_year The page also contains a link *Hot words older than a year*, to which some of the 2020 hot words have been moved.

examples help to harvest relevant headwords (e.g. ***covidi**ot*, ***covid** party*, ***corona**sceptic*, ***corona**viruslike*, *etc*., and *long-hauler*, defined as 'a **COVID**-19 patient who is suffering from [. . .]'). However, the method fails to retrieve words that are not morphologically derived from the patterns and that are related but not specific to the pandemic, i.e. headwords whose defining words do not match such patterns (e.g. no word matches the patterns in the definition and usage examples of *social distancing*). Wiktionary revision logs are the same for all language editions and the number of editions/editors per page can be extracted regardless of any target topic. Details on the processing required to exploit the logs are given in Section 2.1.

2.1 Data processing

The history dump of Wiktionary is a large file released on a regular basis, which contains every version of all articles, stored after each individual contributor's edition. For each revision, the username of the contributor, or the IP address (for unregistered users) is provided, as well as the revision date. The files released on January 1, 2021 were downloaded[3] and processed for the English and French editions of Wiktionary so as to extract, for each month and for each article, the number of revisions and the number of unique contributors.[4] Several pre-processing steps were performed to discard data irrelevant to the present work:

- discussion pages, user talk pages, etc.
- parts of speech other than common noun and proper noun, verb, adjective and adverb.
- pages related only to inflected forms or entries in a language different from the target language (e.g. the page of the English Wiktionary which describes the Vietnamese headword *siêu vi corona* 'coronavirus' was ignored).
- revisions related only to other language sections. For example, the revision dated 5 November 2020 on the *coronavirus* page of the English Wiktionary, that resulted in the addition of the derived term *coronaviraal* to the Dutch language section was ignored.

After discarding irrelevant pages and revisions, more than 14 million revisions remained for the English Wiktionary and more than 27 million for the French language edition.

Studies that focus on Wiktionary (or Wikipedia) revisions differ in whether they take into account the revisions performed by bots and by anonymous users. A bot is

3 https://dumps.wikimedia.org/

4 The computing is an extension of the work done by Sajous et al. (2020) to produce WIND, a resource which contains the dates of inclusion of Wiktionary headwords.

a program devised to automatically perform specific types of revision targeting a range of articles (mainly formatting, importing audio files, etc.). Such automatic editions amount to 45% of the revisions in the English Wiktionary and 62% in the French edition. A contributor to Wiktionary may be identified by a registered account or an IP address. Regular contributors generally create an account while occasional contributors may edit an article "anonymously". Dismissing or taking into account anonymous revisions (which represent 7.4% of the revisions in the English Wiktionary and 4.7% in the French Wiktionary) is a matter of debate. They are discarded in some studies on the grounds that anonymous users are less experienced or trustworthy than registered users and that identifying Internet users by their IP addresses is a rough approximation. Several contributors can indeed use the same IP,[5] while a given contributor can use several IPs,[6] as was the case during the present study (see the discussion on *myroblyte* in Section 3.1). The objection could however apply to registered accounts too: different contributors sharing the same account is probably an exception, but it happens that a single user owns several accounts. Since the present study is concerned with the tendency of articles to be revised many times by possibly many people (not only experienced or reputable Wiktionarians), I was tempted to argue that there is no a priori reason for ignoring anonymous contributions (while revisions performed by bots are not relevant). The quantitative investigations presented in Section 3.1 show that there is no definitive answer as to whether considering or ignoring anonymous contributions is the best option. Regarding qualitative considerations, discarding anonymous contributions poses a risk of overlooking relevant words. For instance, in 2020, the articles for the synonyms *R0*, *basic reproduction number* and *basic reproduction ratio* were created from the same IP address. Whatever their rankings, these words would have gone unnoticed if anonymous contributions had been ignored.

Regarding existing headwords (i.e. those created prior to 2020), it is not difficult to detect new senses added to Wiktionary in 2020 by using the revision log, but the main focus in the present study is on any kind of updates: additions, modifications or replacements of definitions, usage examples, semantic relationships, translations, usage notes, etc. Beyond new meanings, such revisions may indicate the need for article reviews (cf. the examples of *ventilator* in Section 2.4 and *comorbidity*, Section 3.5), which is information that lexicographers may find useful.

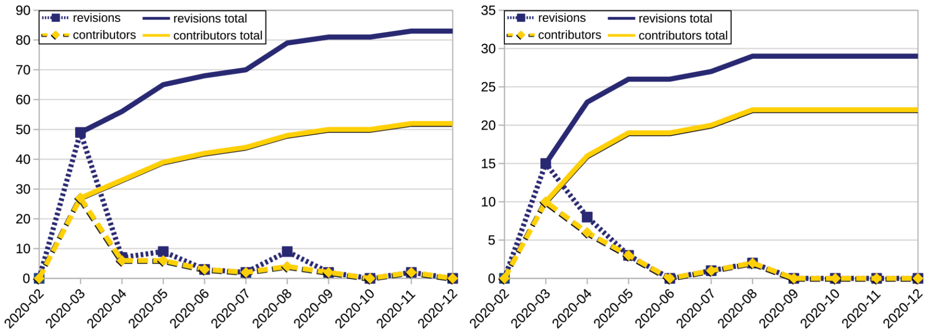
---

5 Especially contributors accessing the Internet from behind an institution firewall.

6 Either intentionally, to deliberately mask one's identity, or unintentionally, as a result of dynamic IP assigning.

2.2 Ranking the new headwords

Looking for the new articles that have the maximum number of revisions and contributors should make it possible to detect headwords related to topical events, notably the Covid-19 pandemic. Figure 1 illustrates the cases of *social distancing* and *flatten the curve*. Dotted lines correspond to the number of revisions/contributors per month, while plain lines represent the total number of revisions/contributors since the creation of the articles.



(a) *social distancing* (b) *flatten the curve*

Figure 1: Number of revisions and contributors for *social distancing* and *flatten the curve*.

The lines for the two words follow the same pattern: a sudden spike of activity when the page is created (typically, when a word comes into usage) and an off-peak period, with occasional contributions (this revision pattern is reminiscent of the pattern described by Renouf (2013) concerning the corpus frequency of *Eyjafjallajökull*, as mentioned in the Introduction). An analysis based on a one-year span is likely to detect the two words *social distancing* and *flatten the curve*, with the former ranking higher (note that the vertical scales in Figure 1 are different). Conversely, words such as, for example, *cognitive bias*, added in April 2020, which received 7 contributions by 4 distinct human contributors are unlikely to be detected. In addition, the two Covid-related words are likely to appear in the first trimester candidate list when performing quarterly analyses.

Wiktionary headwords can be represented in a coordinate system whose axes correspond to the number of revisions and unique contributors of each headword. The 31,107 new headwords added to the English Wiktionary in 2020 are depicted by a scatterplot in Figure 2. The article *COVID-19* was modified 115 times by 63 unique contributors and is therefore represented by the coordinate point (115, 63). A given coordinate point can correspond to several headwords. For example, 17,415 words have not been modified since their creation. They are all represented by the coordinate point (1, 1). A less extreme case, the headwords *self-isolate* and *Wuhan coronavirus*

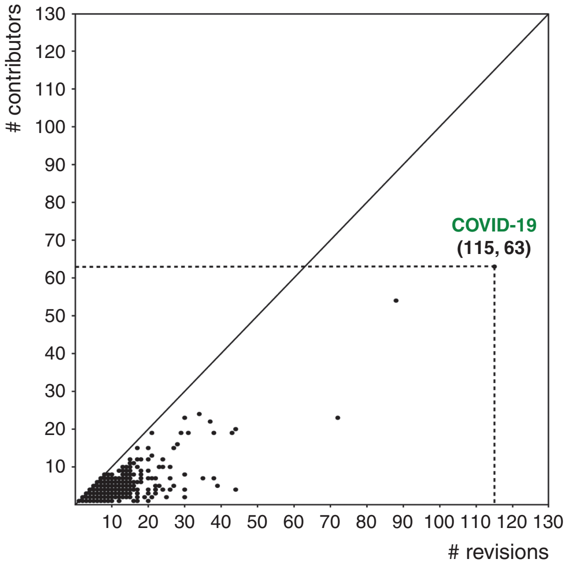


Figure 2: Distribution of the 31,107 lemmas added to the English Wiktionary in 2020.

were each modified 16 times by the same number (11) of distinct contributors. They are therefore represented by the same coordinate point (16, 11). All the points are located along or below the diagonal line (i.e. the *contributors=revisions* line) because there obviously cannot be more contributors than contributions for a given headword. The points along the diagonal line are those for which each revision was made by a distinct contributor. For instance, the words *coronoia*, *Wuhan shake* and *Zoombombing* were modified 8, 7 and 4 times, respectively, each time by a different contributor.

This kind of diagram enables a geometrical interpretation of the headwords' location. The rightmost points of the diagram (i.e. those with the highest abscissa) are those corresponding to the most heavily revised pages. The upmost points (i.e. those with the highest ordinates) are those corresponding to the headwords edited by the greatest diversity of contributors. Given two points having the same abscissa, the upmost point corresponds to the headword revised by a greater diversity of contributors. For example, the two headwords *flatten the curve* and *Medusavirus* 'a virus that infects amoeba' were each modified 30 times in 2020, and have a similar creation date (February and March, respectively). However, the 30 edits of *flatten the curve* were made by 23 distinct contributors, compared to 8 contributors for *Medusavirus*.

Four ranking scores were tested to detect potential Covid-related neologisms. Given a headword *h*, the ranking scores are defined as follows:

1. $revs_h$ = raw number of revisions for *h*
2. $contribs_h$ = raw number of distinct contributors for *h*

3. $dist_h$ = distance from the origin of the coordinate system to the point ($revs_h$, $contribs_h$) $= \sqrt{revs_h{}^2 + contribs_h{}^2}$
4. $prod_h$ = product of the number of revisions and the number of contributors[7] $= revs_h x contribs_h$

The geometrical interpretation of the first three ranking scores is depicted in Figure 3 (the product score has no geometrical interpretation). The revisions-based score orders the headwords from right to left. When two headwords have the same abscissa (i.e. the same number of revisions), they are ordered by their ordinate value (their number of contributors), i.e. the upmost headword is ranked first. For instance, the initially equally ranked (4th position) *Wuhan pneumonia* and *Mount Mayon* (a volcano in the Philippines), whose coordinates are (44, 20) and (44, 4), are finally ranked fourth and fifth. Similarly, the contributors-based score orders the headwords from top to bottom. When two headwords have the same ordinate (i.e. the same number of contributors), they are ordered by their abscissa value (i.e. their number of revisions), i.e. the rightmost headword is ranked first. For instance, the equally ranked (4th position) *myroblyte* (see Section 3.1) and *flatten the curve*, whose coordinates are (72, 23) and (30, 23), are finally ranked fourth and fifth, which, in this case, is not the best option. Finally, the distance-based score orders the headwords according to their remoteness from the origin of the coordinate system.

2.3 Ranking the existing headwords

The scores introduced in Section 2.2, devised to rank Wiktionary new entries, are based on raw numbers of revisions and contributors. Using raw numbers to rank existing entries would not make any sense. We can indeed expect the revision rate of Wiktionary articles to depend on the nature of the entry, i.e. whether it is a frequent or a rare word, polysemous or monosemous, belonging to a specialised field or to the general language (knowing that these characteristics are related). For example, the larger spike observed in 2020 for *coronavirus* when compared to that observed for *virus* in Figure 4(a) is all the more noticeable as the article corresponding to the frequent and polysemous word *virus* is regularly revised, while the entry *coronavirus* rarely is. Another telling example is the number of revisions of *masks*, *facemask* and *surgical mask*, as depicted in Figure 4(b). If we consider the 2020 period globally, the three articles received a similar number of revisions (36, 36 and 34, respectively). However, their “usual” yearly revision values are very different.

7 The product can be normalised to values between 0 and 1 by dividing the score by the maximum number of revisions and the maximum number of contributors. Normalising the product, however, is useless since it does not change the ranking order.

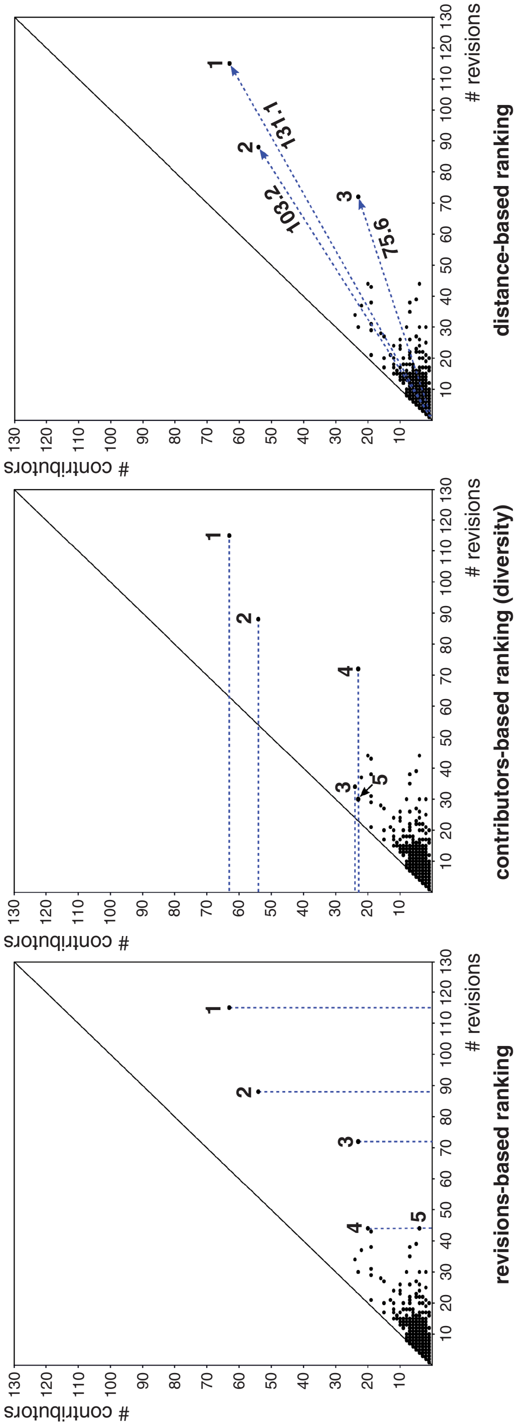
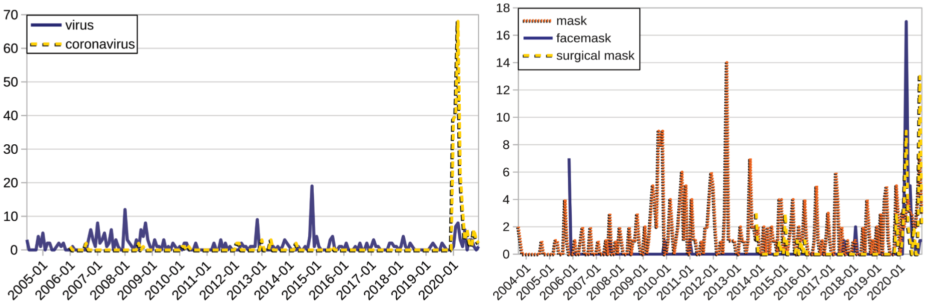


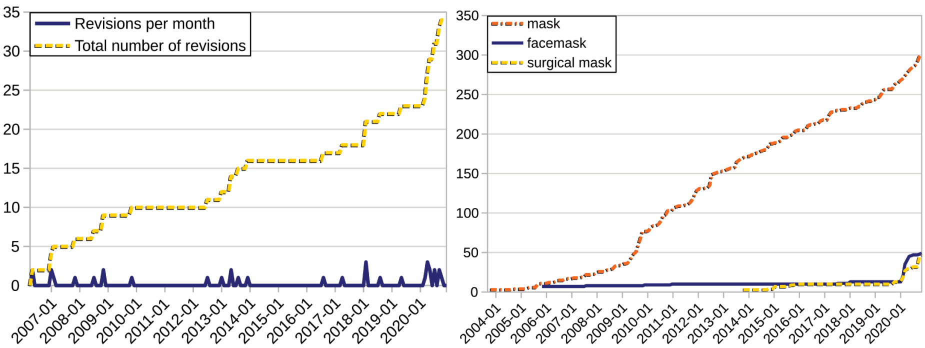
Figure 3: Ranking scores based on revisions, contributors and distance.

Another way to uncover unusual increases in the number of editions is to represent the total number of revisions (or contributors) for a given headword, as depicted in the time-graphs in Figure 5. Figure 5(a) shows that revisions performed over several consecutive months may result in jumps that can be observed for the resulting period. Figure 5(b) shows the total number of revisions for *mask*, *facemask* and *surgical mask*. The increase in the number of revisions is in line with the usual trend for *mask*, while the increases for *facemask* and *surgical mask* are more noticeable.



(a) *virus* and *coronavirus* (b) *mask, facemask* and *surgical mask*

Figure 4: Monthly revision frequencies in the English Wiktionary.



(a) monthly and total revisions for *comorbidity* (b) total number of revisions for *mask, facemask* and *surgical mask*

Figure 5: Monthly and total number of revisions in the English Wiktionary.

The boxplot in Figure 6 statistically confirms these observations: With respective mean values of 3.1, 5.9 and 16.7 (median values of 0, 2 and 15), *facemask* was revised 12 times more than usual, *surgical mask* 5.8 times more and *mask* only 2.2 times more. The two upmost circles in the figure (which represent extreme values) correspond to the 2020 number of revisions for *facemask* and *surgical mask* (another extreme value, observed for *facemask*, correspond to the 7 revisions made

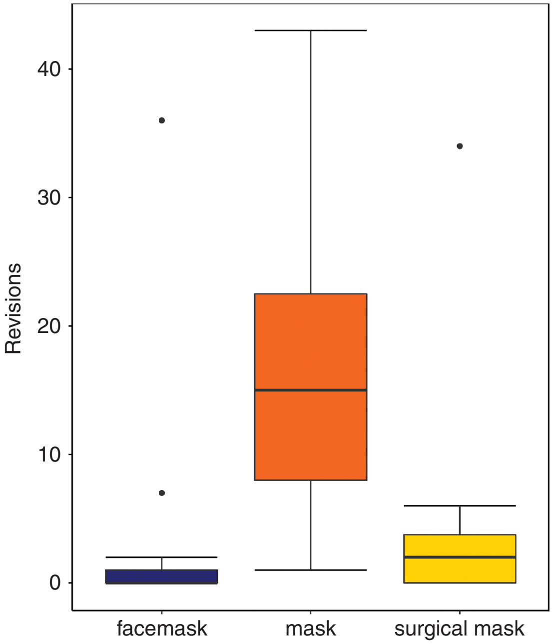


Figure 6: Distribution of the yearly number of revisions in the English Wiktionary.

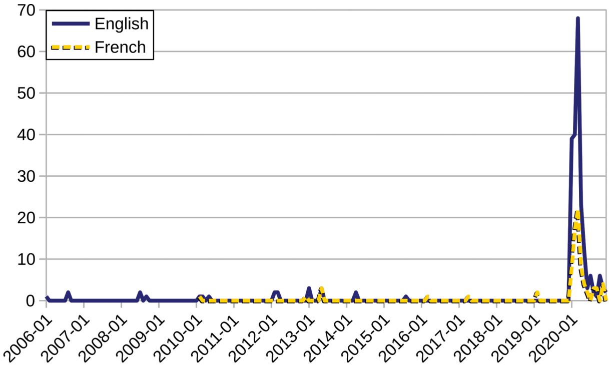


Figure 7: Number of revisions for *coronavirus* in the English and French Wiktionary.

in November 2005 when the article was created). Conversely, the 2020 value for *mask* is not identified as an extreme value.

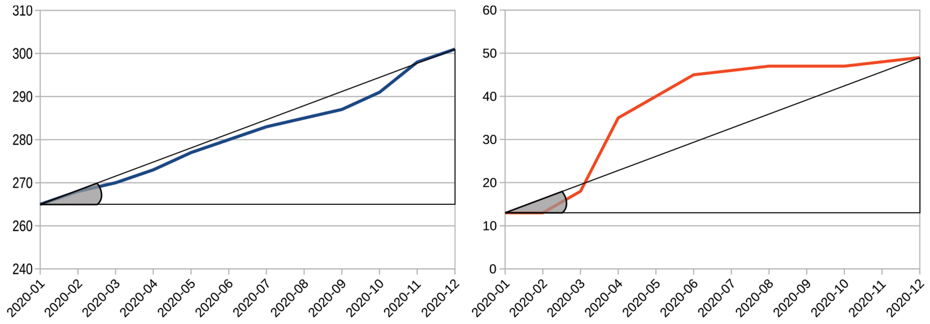
Regardless of the linguistic characteristics of the headwords, the revision rate may differ from one language edition to another. For example, the evolution of the number of revisions for the article *coronavirus* follows a similar trend in the English and French Wiktionary, but with different magnitudes (cf. Figure 7).

Detecting a particular “stir” around a headword is like looking for the extreme values of the boxplot in Figure 6. It therefore requires comparing the number of revisions over a given period to its *usual* revision rate, just as extracting keywords by comparing a focus corpus to a reference corpus requires the use of relative frequencies, not raw frequencies. The scores used to detect the most unusually revised articles compare the number of revisions (or contributors) over a given period to the usual (mean or median) number of revisions (or contributors) over similar time spans. Given a target period p and headword h, the scores are calculated as follows:

1. $avgRevsRatio_p(h) = \frac{1 + revs_p(h)}{1 + avg(revs(h))}$
2. $medianRevsRatio_p(h) = \frac{1 + revs_p(h)}{1 + median(revs(h))}$
3. $avgContribsRatio_p(h) = \frac{1 + contribs_p(h)}{1 + avg(contribs(h))}$
4. $medianContribsRatio_p(h) = \frac{1 + contribs_p(h)}{1 + median(contribs(h))}$

Medians and averages are calculated over the period that spans from the creation of the article corresponding to the headword h to the month before period p begins. A constant (here, 1) is added to the denominator (and to the numerator, for balance) so as to avoid divisions by zero. The median value can be null (as we saw above with *facemask*), but the average value should not be, as all the articles have been edited at least once (when they were created). However, certain revisions (performed by bots or anonymous contributors) are discarded in some of the experiments described below, which makes the addition of a constant necessary.

Ranking the headwords according to the slope of the curve for a given period was tempting. The slope accounts for the increase in the number of revisions (or contributors) over a given time span. For both the English and French language editions, the scores based on slope values performed poorly. As the slope is equal to the ratio between the number of revisions (or contributors) and the length of the time span, its value is proportional to the raw number of revisions (or contributors), and disregards the corresponding *usual* amount, which explains the low results. Figure 8, which consists of two enlargements of Figure 5(b), illustrates the situation for *mask* and *facemask*. Although it is clearly visible in Figure 5(b) that the two words have different usual revision rates, Figure 8 shows that their slope values on the 2020 period are the same. The slope-based score was therefore abandoned and is not further discussed.



(a) *mask* (b) *facemask*

Figure 8: Headwords with similar slope values over the year 2020.

2.4 Annotation of headwords

In order to assess the performances of the ranking scores, several sets of top-ranked headwords were annotated for each metric introduced in Section 2.2 with the binary flag 'related to Covid-19' (yes/no). New headwords were annotated to detect true formal neologisms, or words that already existed but were too rare or too specialised to enter a dictionary before. Existing headwords were annotated to detect semantic neology or articles that potentially deserve an update. The interpretation of the "relatedness to Covid-19" criterion encompasses words whose referents are in a direct or indirect relationship with the virus and the disease, medical care, controlling the spread of the pandemic, statistical analysis, consequences of the pandemic on professional activities and social lives, as well as humorous coinages. For the English and French language editions of Wiktionary, I annotated, for each data source,[8] and each relevant ranking score (cf. Sections 2.2 and 2.3):

- the first 200 new headwords, ranked over the whole 2020 period;
- the first 200 existing headwords, ranked over the whole 2020 period;
- the first 100 new headwords, ranked over each trimester;
- the first 100 existing headwords, ranked over each trimester.

The different (overlapping) sets of headwords represented a total of 3,070 English and 3,168 French words to be annotated. The words were stored in four groups of spreadsheets, setting apart new and existing entries, English and French words. Each word was accompanied by a hyperlink to the online article, along with the definition of the first sense as it stood in Wiktionary. In most cases, the annotation

8 Data sources are discussed in Section 3.1.

was rather self-evident.[9] Conversely, some headwords required further investigation, e.g. reading the definition or looking for additional encyclopaedic knowledge. The definition taken from Wiktionary was intended to help annotate new entries rather than existing, polysemous ones, that require a look at the online article (and, sometimes, at the differences between the versions of articles before/after 2020) or other sources. Encyclopaedic knowledge was especially necessary for annotating words related to the fields of pathology and pharmacology. For example, two drugs related to hydroxychloroquine were positively annotated: *pamaquine* (existing in the English Wiktionary since 2009) and *quinium* (added to the French Wiktionary in 2020). Revisions of entries denoting other drugs may have been motivated by pandemic-related reasons, such as drugs used in the treatment of respiratory diseases (e.g. *bambuterol*, used in the treatment of asthma). However, in the absence of clear evidence of a relation with Covid-19, such headwords were negatively annotated. On the same lines, *extractor fan* may be related to air purification that helps prevent the spread of the virus. However, this entry, created in June 2019, makes no reference to such a meaning. It was deemed too general and was therefore negatively annotated. Conversely, *ventilator* was annotated positively. Although not related to the Covid pandemic when used as a synonym of *fan*, the 2020 updates clearly target the *medical ventilator* sense. The previous synonymic definition '(medicine) A respirator' was changed to '(medicine) A machine that moves breathable air into and out of the lungs of a patient who is unable to breathe sufficiently', with *respirator* now appearing as a hypernym. A picture of a medical ventilator has been added, as well as the derived term *tank ventilator* and numerous translations.

In the case of French borrowings from English, the prior annotation of the English word helped. For instance, it would have been hard to come to a decision in the case of the neologism *doomscrolling* '(informal) The practice of continually reading Internet news about catastrophic events' retrieved from the English Wiktionary by only reading its definition. The definition may refer to one's state of astonishment when following the news after the pandemic outbreak or when the first lockdown was decided. But *catastrophic events* can, sadly, designate numerous other facts. The problem was finally easily solved, due to the presence of the *Coronavirus* category at the bottom of the article. In the French Wiktionary, the article *doomscrolling*, which mentions the borrowing from English, but does not mention the pandemic in the definition or the usage examples, is devoid of any topical category. The previous annotation of the English word led to a positive annotation. In the French Wiktionary, the Anglicism *contact tracing* was added in April 2020, and its annotation did not raise any difficulty. Conversely, *tracking* was debatable. Until

9 Given the number of regionalisms, occasionalisms and dated words, in addition to words of different subcultures, a quick look at the definition was necessary, even for French words. *Self-evident* therefore means non-ambiguous here, rather than *immediate*.

2020, the corresponding article described the English gerund. In April 2020, the description of the French Anglicism *tracking* was added and defined as *surveillance de masse des populations par pistage de tous les citoyens* 'mass surveillance of populations by tracking all citizens'. Though the definition does not refer to controlling the spread of the pandemic, the three usage examples are related to this goal, in particular to the use of cell phones for contact tracing which was then a matter of debate, echoing discussions on other freedom-destroying laws. Knowledge of current events helped annotate the headword positively. In the English Wiktionary, *pastette* was only described as the plural form of the Italian *pastetta*, a variant of *pastella* 'batter' until 2020. The English noun was added in 2020 and described as a synonym of '*Pasteur pipette*'. Although the use of this instrument is not specific to blood sample collection for Covid-19 testing, the addition of the entry to the dictionary is obviously related to the pandemic and the headword was therefore positively annotated.

Some additions and some words already in the dictionary refer to things of the past. For example, *méthode Raspail* entered the French Wiktionary in March 2020 and refers to a hygiene system named after its creator François Vincent Raspail, mainly based on handwashing and dating back to the nineteenth century. Despite the lack of exclusive connection to Covid-19, the 2020 addition of this old preventive measure, simultaneously with the revisions of *gel hydroalcoolique* 'alcogel' and the addition of *geste barrière* 'practice intended to avoid the spread of a virus' argued in favour of a positive annotation. Although the 7 revisions of *quarantine flag* (in the nautical field, the flag that was hoisted by a ship to signal that it had contagious disease aboard) by 5 distinct contributors to the English Wiktionary, which resulted in a rewording of the definition and the addition of four translations and a reference, are striking, the word was deemed too indirectly related to the pandemic to be assigned a positive annotation (the flag is said to have been *formerly* hoisted and the reference dates back to 1916).

Lastly, some words were close to being given a positive annotation they did not deserve. With the videoconferencing software in mind, it was tempting to annotate positively the French *zoom* and *zoomer*, and the English *zoomer*, without checking the corresponding definitions. However, the French words are only related to the camera lens and the revisions of the English *zoomer* are related to the generational designation (*active boomer*, *member of Generation Z*).[10] The British slang *lurgy*, denoting a fictitious, or uncategorised, infectious disease with cold or flu-like symptoms, that renders one unable to work, was a good candidate. However, several occurrences found in newspaper articles dating back to Fall 2019 (i.e. *before times*), whose topic

10 Derivatives of *Zoom* (the videoconferencing software) retain the initial upper case in English, according to the English Wiktionary.

was the ironically named season "of dreaded lurgy" in reference to people seeking excuses for work absenteeism, led to the word being rejected.

The scatterplot in Figure 9 depicts, for the English and the French Wiktionary, the Covid-related words in blue circles and the negatively annotated words in yellow circles. Grey triangles correspond to the superposition of several words, some of which are related and others not related. Empty circles close to the origin of the coordinate system correspond to words that were not annotated because they did not rank high enough in any configuration.

Simple linear regressions were performed for the two language editions and the red lines represent the models fitting the distributions. A first observation is that the points corresponding to positively annotated headwords are mostly above the regression line. Given several articles that received the same number of revisions, the articles related to the Covid pandemic are those edited by the largest number of contributors. This finding is further investigated in Section 3.2.

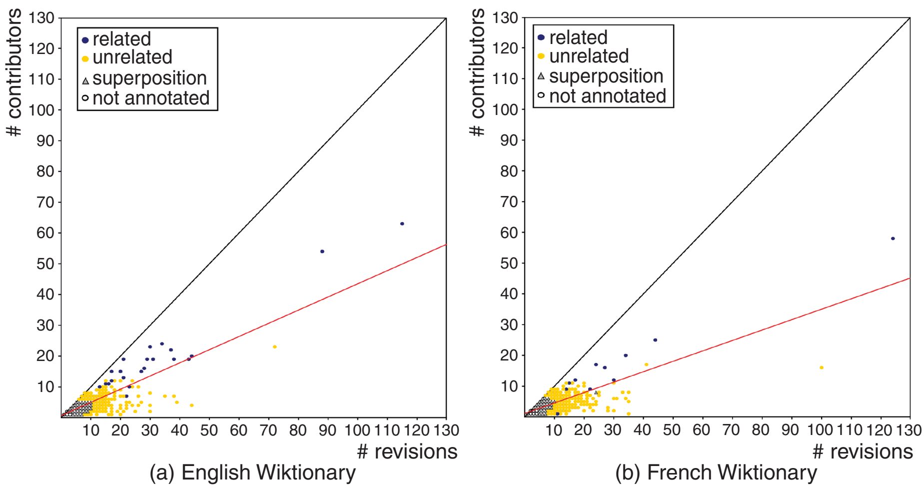


Figure 9: Distribution of the headwords added to Wiktionary in 2020 with respect to revisions and contributors.

3 Results

3.1 Contributor types

The performances of the ranking scores were calculated from different data sources in order to evaluate which kinds of revisions were worth taking into account with respect to the contributor types (cf. Section 2.1). Figures 10(a) to 10(d) show the results

obtained when considering all revisions compared to the results obtained when ignoring revisions performed by bots and/or by anonymous contributors. Figures 10(a) and 10(b) correspond to the English Wiktionary while Figures 10(c) and 10(d) correspond to the French Wiktionary. For the two language editions, the results obtained with the ranking score based on the number of revisions are visible on the left-hand side, i.e. in Figures 10(a) and 10(c). The results obtained with the ranking score based on the number of contributors are visible on the right-hand side, i.e. in Figures 10(b) and 10(d). The line chart shows the percentage of new entries related to the pandemic on the ordinate, as a function of the number of candidates examined (ranging from 1 to 200), on the abscissa.

Regardless of the data sources, the results are better for the English Wiktionary than for the French edition and the ranking score based on the number of contributors performs better than that based on the number of revisions. These observations are further discussed in Section 3.2. The experiments confirm that discarding the revisions performed by bots generally improves the results. Regarding anonymous contributions, conclusions are mitigated. Discarding these contributions significantly lowers the results obtained with the French Wiktionary. It also lowers the results obtained with the English Wiktionary when using the ranking score based on the number of revisions, but it improves those based on the number of contributors. In Figure 10(b), a clear advantage is visible in the top of the list, up to rank 18. The first downshift observed in this figure for the data involving anonymous contributors is due to the headword *myroblyte* ‘a saint whose relics or place of burial produce or are said to have produced the Oil of Saints’. With 71 revisions coming from 22 IP addresses, the headword reaches the third rank. A closer look revealed that the addresses were most likely assigned to the same machine.[11]

Given that the ranking score based on the number of contributors outperforms the ranking score based on the number of revisions whoever the authors of the revisions, the experiments described in Section 3.2 were performed with the source of data that produced the best results with the contributors-based score, for each language edition, i.e. the “no bots, no anonymous” option for the English Wiktionary and “no bots, with anonymous” for the French edition. The best choice regarding data sources, however, is unstable. The experiments conducted for each trimester led to better results for the English Wiktionary when the anonymous contributions were taken into account. The results presented in Section 3.3 were produced with the “no bots, with anonymous” option.

11 All of them have the same two left numbers, and are probably due to dynamic IP assigning. A comparable number of revisions stemming from the same addresses is observed for the same headword in the French Wiktionary.

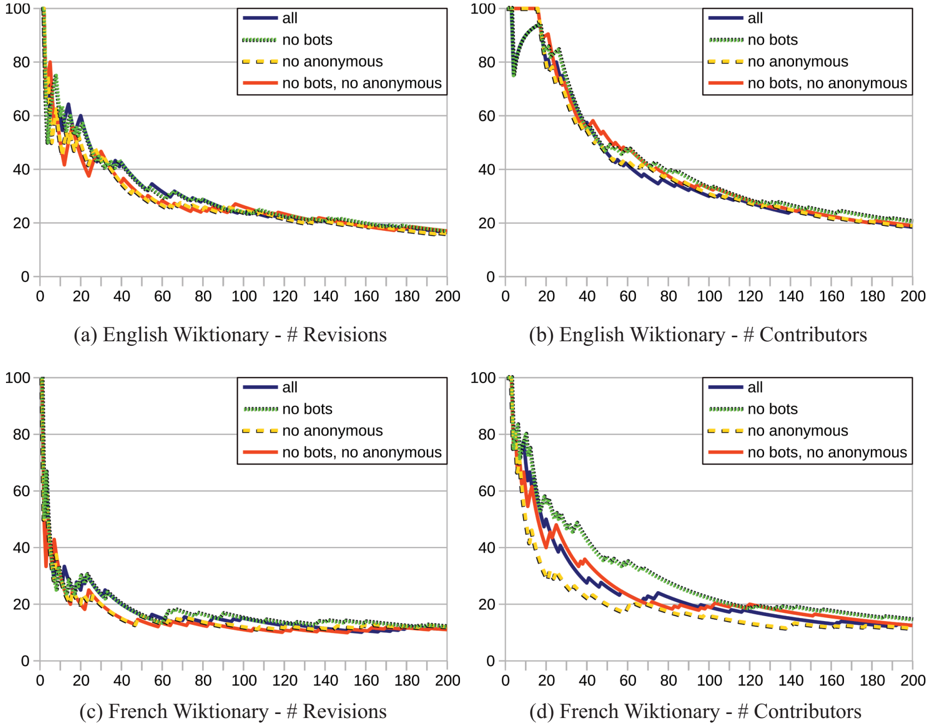


Figure 10: Influence of the contributor types on the ranking scores.

3.2 Yearly ranking of the new headwords

The more up-to-date a dictionary is, and the more exhaustive its list of headwords, the more likely a new headword is to be a neologism. The top-ranked new additions to Wiktionary, according to the metrics introduced in Section 2.2, were therefore inspected to detect formal neologisms. Table 1 reports, for the English edition of Wiktionary, the 20 most heavily edited new entries in 2020, sorted by number of revisions, number of unique contributors, and by the two combinations (product and distance) introduced in Section 2. Grey cells indicate headwords that are not related to the pandemic, while the others are.

When ordered by number of revisions, less than half of the top-ranked entries (9 out of 20) are related to the pandemic. When ordered by number of contributors, 90% of them (18 out of 20) are positively annotated. This result seems to confirm the initial hypothesis: the most frequently edited Wiktionary pages, especially pages edited by many distinct contributors, can help detect topical neologisms.

Looking down the list after the 20th rank of the contributors-based score helps detect the following relevant words:[12] *fever clinic* (21), *self-isolate* (27), *coronoia* (29), *before times* (38), *doomscrolling* (41), *Wuhan shake* (42), *maskne* (43), *SARS-CoV* (48), *rat-licker* (54), *contact trace* (58), *maskhole* (61), *elbow bump* (74), *mascne* (88), *plandemic* (94), *China virus* (96), *Covidtide* (103), *corona virus* (134), *case fatality rate* (144), *coronasceptic* (180), *elbow shake* (182), *antimasker* (238), *covid-19 party* (250), *long-hauler* (272), *corona belly* (484), *community spread* (492), etc.

Table 1: 20 most frequently edited pages in the English Wiktionary 2020 additions, according to different ranking scores (data source: no bots, no anonymous).

Rank	# Revisions	# Contributors	Product	Distance
1	COVID-19	COVID-19	COVID-19	COVID-19
2	social distancing	social distancing	social distancing	social distancing
3	Mount Mayon	covidiot	Wuhan pneumonia	Mount Mayon
4	Wuhan pneumonia	flatten the curve	Wuhan virus	Wuhan pneumonia
5	Wuhan virus	Wuhan virus	covidiot	Wuhan virus
6	Bicol Region	Wuhan pneumonia	flatten the curve	Chinese virus
7	Peja	covid	Chinese virus	Bicol Region
8	Chinese virus	social distance	covid	Peja
9	Berat	SARS-Co V-2	infectious disease specialist	covidiot
10	Medusavirus	Chinese virus	social distance	flatten the curve
11	Vlorë	infectious disease specialist	SARS-CoV-2	Berat
12	Kizilsu	contact tracing	contact tracing	Medusavirus
13	covidiot	self-isolation	Kung Flu	covid
14	infectious disease specialist	Kung Flu	Wuhan flu	infectious disease specialist
15	flatten the curve	Wuhan flu	Medusavirus	Vlorë

12 Whether such neologisms should be added to a dictionary headword list is not the focus of the present research and depends on the editorial policy. The final decision is up to the lexicographer, and is not discussed here.

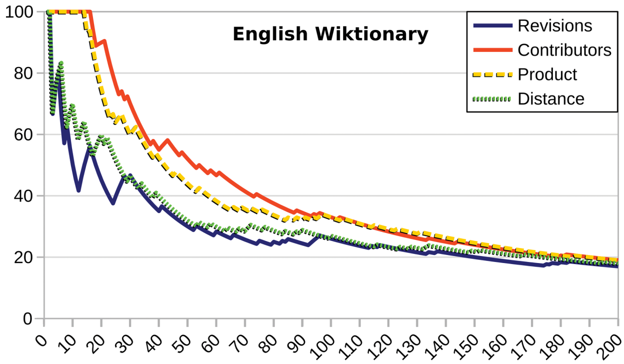
Table 1 (continued)

Rank	# Revisions	# Contributors	Product	Distance
16	covid	COVID	self-isolation	Kizilsu
17	world map	Spleef	world map	world map
18	Accompong	Trumpster fire	Bicol Region	social distance
19	Arlberg	self-quarantine	Mount Mayon	SARS-CoV-2
20	sinoatrial node	Wuhan coronavirus	wokefish	contact tracing

The same ranking score applied to the French Wiktionary provides the following Covid-related words: *Covid-19* (1), *covid* (2), *COVID-19* (3), *covidiot* (5), *déconfinement* ‘deconfinement, lockdown removal’ (6), *distanciation sociale* ‘social distancing’ (8), *covidé* ‘sick from Covid-19’ (9), *déconfiner* ‘deconfine, remove lockdown’ (10), *reconfinement* ‘reconfinement, new lockdown’ (12), *masque chirurgical* ‘surgical mask’ (18), *coronavirus 2 du syndrome respiratoire aigu sévère* ‘SARS-CoV-2’ (19), *méthode Raspail* ‘hygiene system based on handwashing’ (21), *télétravaillable* ‘(work) that can be done by teleworking’ (25), *covidien* ‘related to, or sick from Covid-19’ (27), *cas contact* ‘contact case’ (31), *distanciation physique* ‘physical distancing’ (34), *gel hydroalcoolique* ‘alcogel’ (35), *doomscrolling* (49), *infodémie* ‘infodemics’ (53), *hydroxychloroquine* (58), *Covid* (60), *pneumonie de Wuhan* ‘Wuhan pneumonia’ (67), *syndémie* ‘syndemic’ (121), *antimasque* ‘antimask’ (124), *démerdentiel* ‘(informal) activity performed with the means available’ (133), *coronasceptique* ‘coronasceptic, who denies the reality or the aftermath of the coronavirus’ (136), *autoconfinement* ‘self-isolation’ (142), *Covid positif* ‘Covid positive’ (156), *coronapiste* ‘temporary cycle lane built during the Covid-19 pandemic’ (186), *raoultiste* ‘supporter of Pr. Raoult’ (214), *candidat-vaccin* ‘vaccine candidate’ (308), *coronavirussé* ‘sick from Covid-19’ (361), *tempête immunitaire* ‘cytokine storm’ (432), etc.

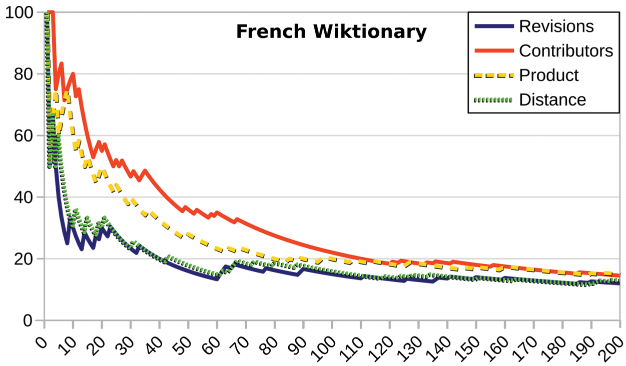
The performances of the different ranking scores are further illustrated in Figures 11 and 12 for the English and French language editions. For both languages and for the four ranking scores, the line charts are similar to those in Section 3.1 and show the percentage of new headwords that are related to the pandemic on the ordinate, as a function of the number of candidates considered (ranging from 1 to 200) on the abscissa.

The same observation can be made for both language editions: the ranking based on the number of unique contributors performs markedly better than the ranking based on the number of revisions. The number of contributors alone even outperforms the combinations of the two measures (with the “product” score only slightly improving the results locally from ranks 115 to 179 and equalling the results of the contributors-based score from rank 187 onwards).



	Ranking score			
Rank	Revs	Contribs	Prod	Dist
5	80	**100**	**100**	80
10	50	**100**	**100**	70
25	40	**76**	64	52
50	30	**52**	44	32
100	26	**33**	**33**	27
200	17	**19**	**19**	18

Figure 11: Performance of the ranking scores for the English Wiktionary new headwords.



	Ranking score			
Rank	Revs	Contribs	Prod	Dist
5	40	**80**	60	60
10	30	**80**	60	60
25	28	**52**	30	30
50	16	**36**	28	18
100	15	**22**	20	16
200	12	14.5	**15**	13

Figure 12: Performance of the ranking scores for the French Wiktionary new headwords.

In order to explain the lower results obtained with the French language edition, a simple linear regression was performed, as depicted in Figure 9 (Section 2.4), where the regression lines appear in blue. Linear regressions are usually performed to confirm that two variables are significantly related. In the case of the number of revisions and contributors, we already know that this is the case, but we are interested in the regression coefficients. With a value of 0.43,[13] the slope of the regression line for the English Wiktionary is greater than that for the French edition (slope value of 0.34).[14] This means that, given two articles selected at random in the English and French editions, that have the same number of revisions, the article from the English Wiktionary is likely to have been modified by a greater number of contributors than that in the French Wiktionary. This finding, combined with the better results obtained for the English language edition, is an argument in favour of the relevance of the “diversity”

13 $F(1, 31104) = 6.64e^{+4}$, p-value < 0.001.
14 $F(1, 11619) = 1.545e^{+4}$, p-value < 0.001.

measure. To go further in this direction, the coordinates of the positively and negatively annotated headwords originating from the English Wiktionary were set apart in two distinct scatterplots, as shown in Figure 13.

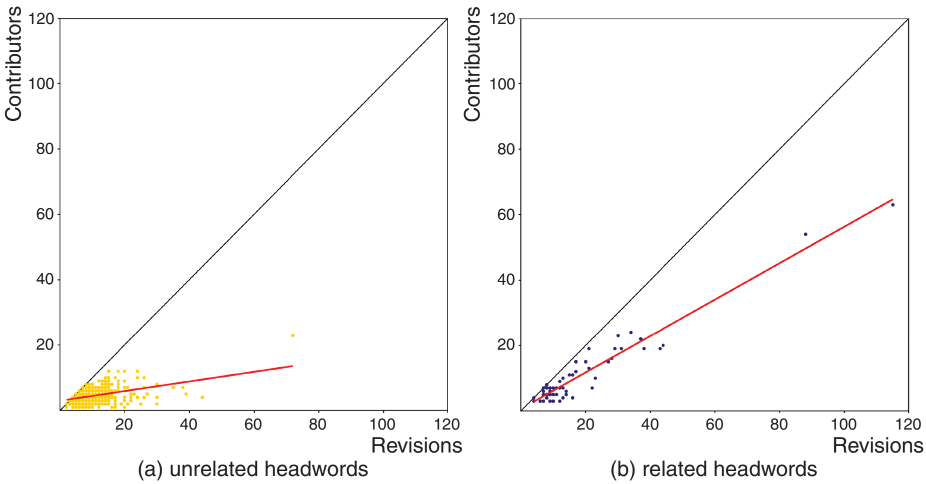


Figure 13: Distribution and regression lines for the related vs unrelated English headwords.

For each distribution, a simple linear regression was performed. The slope coefficients are 0.56[15] for the Covid-related words and 0.15[16] for the words that are not related (regression lines are depicted in red in Figure 13). This means that, given a number of revisions, a positively annotated headword is likely to have been revised by a larger number of contributors than a negatively annotated headword with the same number of revisions.

For each headword of the two annotation sets, the ratio between the number of contributors and the number of revisions was calculated. This ratio can be understood as the "local" diversity for individual headwords: diversity(h) = contributors(h) / revisions(h). Its maximum value is 1 when all the revisions were made by distinct contributors, i.e. when contributors(h) = revisions(h). The ratio is low when all the revisions were made by the same contributor, and especially when the number of revisions is high. The boxplots in Figure 14 represent variations of this ratio. Figure 14(a) shows the difference in diversity between related and unrelated headwords in the English Wiktionary. The median value is 0.62 for the positively annotated words and 0.56 for the words annotated negatively. A Welch two-sample t-test shows that the difference

15 $F(1, 51) = 730.2$, p-value < 0.001.
16 $F(1, 633) = 119.7$, p-value < 0.001.

is statistically significant.[17] The same experiment was conducted on the French Wiktionary. For this language edition, the diversity is also higher for the positively annotated words than for words annotated negatively. This time, however, the difference was not statistically significant.

To conclude on the importance of diversity, we compared the diversity ratio for the 1598 annotated new headwords (688 originating from the English Wiktionary and 910 from the French Wiktionary), regardless of the annotation. The variation in diversity is depicted in Figure 14(b). The diversity is greater in the English Wiktionary (median value of 0.57 compared to 0.5 for the French Wiktionary) and the difference is statistically significant.[18] Once again, this result, together with the lower performances observed for the French Wiktionary, drives home the importance of the diversity measure.

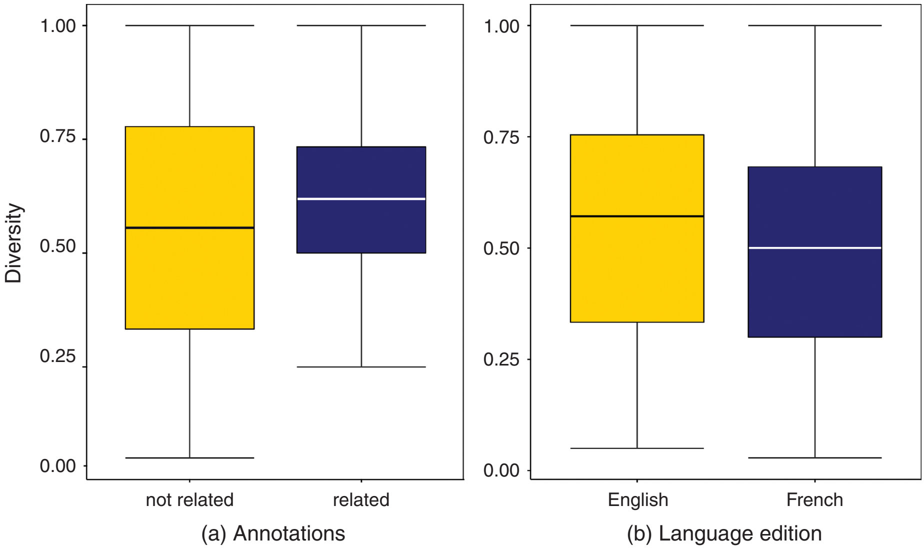


Figure 14: Variation in the diversity ratio depending on annotations and the language editions.

3.3 Quarterly ranking of the new headwords

The method proposed above is based on the analysis of Wiktionary revisions on a whole year basis. However, retrospectively identifying neologisms one year after the Covid-19 outbreak (and after lists of neologisms have proliferated) might seem like making a weather forecast for the day before. For dictionaries such as the

17 $t(71) = -2.0905$, p-value < 0.05.

18 Welch two-sample t-test: $t(1423) = 3.7912$, p-value < 0.001.

French *Petit Robert* and *Petit Larousse*, which are updated once a year, the method based on yearly analyses makes sense.[19] However, some online dictionaries are updated more or less continuously and, among them, the OED is ordinarily updated quarterly. In order to assess the validity of the proposed method for updating such dictionaries under conditions closer to reality, I examined what would have been the top-ranked entries by the end of the four 2020 trimesters (thereby mimicking the quarterly updates that usually occur in the OED). The number of pandemic-related neologisms that would have been detected among the first 100 candidate headwords by the end of each trimester, according to the contributors-based score, is reported in Table 2.

Table 2: Number of Covid-related additions, depending on the number of candidates inspected.

# candidates	English Wiktionary				French Wiktionary			
	T1	T2	T3	T4	T1	T2	T3	T4
5	4	5	2	2	2	4	0	3
10	8	9	4	3	3	6	1	4
25	12	12	12	6	3	9	4	6
50	18	14	14	8	3	11	6	8
100	20	18	16	13	4	14	9	16

The relevant words retrieved from the English Wiktionary are listed below. The headwords in regular type font are those which were already detected during the previous trimesters while headwords in boldface indicate previously undetected words:[20]

- **T1: social distancing**, **COVID-19**, **Wuhan pneumonia**, **flatten the curve**, **Wuhan coronavirus**, **SARS-CoV-2**, **Kung Flu**, **Wuhan virus**, **COVID**, **social distance**, **Wuhan flu**, **SARS-CoV**, **case fatality rate**, **Chinese virus**, **self-isolate**, **covid**, **community spread**, **self-isolation**, **self-quarantine**, **noncoronaviral**
- **T2:** COVID-19, social distancing, **infectious disease specialist**, **covidiot**, SARS-CoV-2, flatten the curve, Wuhan virus, self-isolation, COVID, Chinese virus, social distance, **corona virus**, self-quarantine, **elbow shake**, Kung Flu, SARS-CoV, **Rona**
- **T3:** COVID-19, covid, social distancing, **maskne**, **mascne**, **plandemic**, Chinese virus, **contact tracing**, **covid-19 party**, covidiot, **doomscrolling**, Wuhan shake, coronoia, **antimasker**, social distance, SARS-CoV-2

19 No Covid-related neologism was added to the printed *Petit Robert* in 2020 (i.e. to the 2021 edition) but some words were added to the online version, which, for the first time, became out of sync with the paper version. See: https://orthogrenoble.net/mots-nouveaux-dictionnaires/entrees-petit-robert-2021/ (last access: 1 June 2021).

20 Discarding previously detected words and upshifting words of lower ranks only results in the addition of *coronapocalypse* at the end of the list of the second trimester.

- **T4: fever clinic**, **rat-licker**, COVID-19, **before times**, China virus, **Covidtide**, Wuhan flu, **coronasceptic**, **long-hauler**, covidiot, Wuhan virus, **long Covid**, **long covid**

In the French Wiktionary, the relevant words detected are the followings:
- **T1: Covid-19, COVID-19, distanciation sociale, pneumonie de Wuhan**
- **T2:** Covid-19, **déconfinement**, **covidiot**, **covid**, **covidé**, **méthode Raspail**, **infodémie**, **déconfiner**, distanciation sociale, **coronavirus 2 du syndrome respiratoire aigu sévère**, **gel hydroalcoolique**, COVID-19, **distanciation physique**, **masque chirurgical**
- **T3: doomscrolling**, covid, **démerdentiel**, COVID-19, Covid-19, **coronapiste**, **taux d'incidence** 'incidence rate', **coronavirussé**, déconfiner
- **T4:** Covid-19, covid, **télétravaillable**, **reconfinement**, **syndémie**, covidiot, déconfinement, **candidat-vaccin**, **cas contact**, coronavirus 2 du syndrome respiratoire aigu sévère, **infectivité** 'infectivity', **antivaccinisme** 'antivaccinism, opposition to vaccination', **télétravaillabilité** 'ability (of a work) to be done by teleworking', **antimasque**, **covidien**, **coronasceptique**

Some of the words detected are true formal neologisms (e.g. *COVID-19*, *Wuhan flu*). Conversely, numerous new headwords already existed before their addition to the Wiktionary list of headwords, as happens in commercial dictionaries. For example, *case fatality rate* has been mentioned by the Office québécois de la langue française in the terminological record *taux de létalité* of its *Grand Dictionnaire Terminologique* since 2009,[21] and probably had long been used by specialists of epidemiology and statistics before it was inventoried in the term bank. The sudden spread of the word in the general language, due to the – no less sudden – spread of the pandemic, motivated the creation of the corresponding article.

Reviewing the lists produced by varying the ranking scores and the data sources is a good idea, as it retrieves headwords that the "globally better" configuration misses. For example, with 6 revisions performed by only 2 registered contributors in the third trimester, *supercontaminateur* 'superspreader' does not rank high enough to be detected by the contributors-based score but ranks 19th with the revisions-based score applied to the "no bots, no anonymous" dataset.

3.4 Yearly ranking of the existing headwords

The same experiments were conducted for existing headwords as for new headwords, by varying the source of data and the ranking scores. Only the main results are reproduced here, while others are summarised.

21 http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=100408 (last access: 10 June 2022).

Just as for the analysis related to new headwords, discarding anonymous contributions slightly improved the results for the English dictionary (the "no bots, no anonymous" option was therefore chosen). As for the new headwords, the scores based on the number of contributors provide better rankings than those based on the number of revisions in the English Wiktionary, as shown in Figure 15. Regarding median and mean values, using one or the other alternatively improves or lowers the results locally. The ratio using the average number of contributors reaches the best result, with 18% of the 100 best-ranked headwords being related to Covid: *coronavirus, hydroxychloroquine, rona, lockdown, superspreader, surgical mask, pandemic, facemask, corona, herd immunity, MERS, MERS-CoV, Coronavirus, Zoom, ventilator, chloroquine, facial mask, SARS*. The other ranking scores produce the same words (but fewer) in different orders. They provide only two extra words – *respirator* and *syndemic* – that are further down in the list (ranks 191 and 1353, respectively).

In the French Wiktionary, considering or discarding anonymous contributions provides quantitatively similar results, with the set of relevant retrieved words differing according to the score used. The contributors-based score performs better, but the difference with the score based on the revisions is less pronounced than when experimenting with the English Wiktionary. Overall, the proportion of relevant words identified in the 100 best-ranked words is lower in the French Wiktionary. The best configuration (ratio involving median values of contributors and no anonymous contributions), which reaches 9%, is half of that obtained in the English Wiktionary. This configuration made it possible to retrieve the words *confinement, coronavirus, pandémie, chloroquine, cluster, télétravail* 'teleworking', *distanciation, contagiosité* and *présentiel* 'in-person activity'. Other configurations retrieved three additional words: *SRAS* 'SARS', *quatorzaine* 'two weeks quarantine' and *corona*.

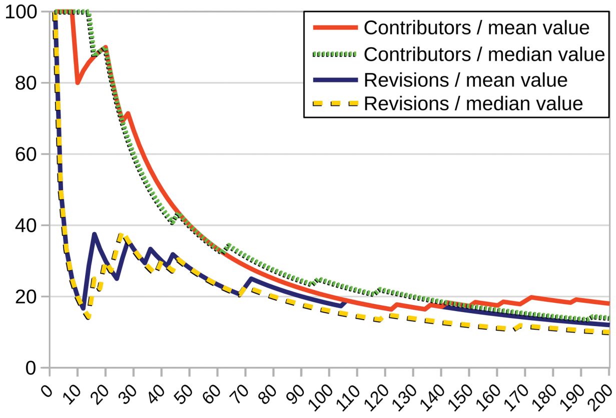


Figure 15: Percentage of Covid-related existing headwords in the English Wiktionary, depending on the ranking scores.

3.5 Quarterly ranking of the existing headwords

For the two languages, discarding the activity of bots improved the results but discarding anonymous contributions barely improved them. Scores based on the number of contributors produced, again, better rankings than those based on revisions, which, sometimes, help detect few extra relevant words. The numbers of relevant existing headwords retrieved from the English and French Wiktionary, as a function of the number of candidates examined, are given in Table 3.

Table 3: Number of identified existing words related to Covid-19, as a function of the number of candidates inspected.

# candidates	English Wiktionary				French Wiktionary			
	T1	T2	T3	T4	T1	T2	T3	T4
5	5	5	0	2	2	1	1	0
10	6	8	0	3	3	2	2	0
25	8	9	2	5	4	4	2	1
50	11	10	2	5	6	8	2	3
100	14	15	3	7	7	9	3	5

Once again, the method achieved better results with the English Wiktionary. For this language edition, the score based on the current/usual contributors ratio (using mean values) performed best. The true positives retrieved by this score applied each trimester to the “no bots, no anonymous” data source are:

- **T1: coronavirus, pandemic, lockdown, corona, Coronavirus, quarantine, Wuhan, rona, herd immunity, panic buying, respirator, superspreader, MERS-CoV, ventilator**
- **T2:** coronavirus, corona, lockdown, **facemask**, rona, **hydroxychloroquine**, **surgical mask**, **Corona**, herd immunity, **pulmonologist**, pandemic, **Zoom**, superspreader, **quar**, MERS-CoV
- **T3:** coronavirus, lockdown, **intensive care unit**
- **T4:** hydroxychloroquine, coronavirus, superspreader, pandemic, **virulent**, **chloroquine**, rona

Varying the data source (i.e. retaining anonymous contributions) retrieved the additional *severe acute respiratory syndrome* (T1) and *viral load* (T3). Changing from mean to median (still with the contributors ratio) retrieved *disinfection* (T1) and *coronary* (T2) while switching to the revisions-based score added *antigen* (T1), *pulmonology* (T2), *immunology* (T3) and *syndemic* (T4).

The same ranking score (contributors ratio, mean value) applied to the same type of revisions (“no bots, no anonymous” option) retrieved the following words from the French Wiktionary:

- **T1: coronavirus**, **confinement**, **chloroquine**, **grippe espagnole** ‘Spanish influenza’, **pandémie**, **SRAS**, **cluster**
- **T2:** confinement, coronavirus, **vidéoconférence**, **pangolin**, **masque** ‘mask’, pandémie, **corona**, **skypero** ‘online apéritif using Skype’, **quatorzaine**
- **T3:** coronavirus, **présentiel**, **jauge** ‘capacity’
- **T4: vaccin, antivaccin, vaccinal, distanciation, écouvillon** ‘swab’

Changing from mean to median only added *stop and go* (T4), while using the revisions-based score additionally produced *virus* (T1), *télétravail* ‘teleworking’ (T2) and *infectiosité* ‘infectivity’ (T4).

The better results observed for the English Wiktionary are related to the greater number of revisions/contributors for some articles. The variation observed for comparable words (e.g. translational equivalents that may have comparable frequencies, degrees of polysemy and specialisation) can simply be due to the number of contributors on the lookout. Another explanation is the possible different degrees of completeness of the articles. For example, the existing headword *comorbidity* ranks high in the second trimester for having been revised several times, as shown in Figure 5(a) (Section 2.3). Although the article was quite up-to-date (the definitions were not modified in 2020), a usage example was replaced by another, more recent, one with an explicit reference to the coronavirus. A pronunciation, an alternative form (*co-morbidity*), synonyms and related words were added, as well as numerous translations. In French, no alternative form exists and the pronunciation was already mentioned in the article before the pandemic. The only revision in 2020 (addition of a recent citation) did not enable the word to be detected.

3.6 False negatives

All the experiments above demonstrated that the proposed method uncovers relevant neologisms, or indicates entries that possibly deserve a review. What the experiments do not say, however, is what the method missed. I therefore examined the ranking of the words included in the two OED updates dedicated to Covid-19 vocabulary. The words added to the OED in April were generally at the top of the list of the headwords detected in Wiktionary, e.g. *Covid-19, social distancing, flatten the curve, Covid, self-isolation, contact tracing, self-quarantine, self-isolate*. Most of the words added in July were already in Wiktionary before 2020. Some of these existing headwords rank quite high, e.g. *corona, surgical mask, MERS, Zoom, dexamethasone, comorbidity*. Others rank much lower down, meaning that the articles received very few revisions, either because they were overlooked by contributors, or

because they were already up-to-date. For example, once *Kawasaki's* is defined as 'synonym of Kawasaki disease', with a hyperlink to the corresponding article, there is not very much left to say. The most noticeable words that the method failed to detect are *shelter-in-place*, added to Wiktionary as a run-on entry under *shelter* in 2020,[22] and words whose ranking scores are very low, e.g. *R0* (dated March 2020), which received 5 revisions made by 3 contributors and *frontliner*, unmodified since its creation in April 2020. Other undetected words or words missing from the nomenclature are those having more popular derivatives (e.g. *contact tracer* ranks relatively low while *contact tracing* ranks high on the list) or equivalents (*community transmission* is absent from Wiktionary, but *community spread* ranks high).

Undetected words are those having too few revisions or contributors. One could speculate that these words are too rare or too specialised to catch contributors' attention. Another variable is Wiktionarians' idiosyncratic contribution patterns. Some contributors make numerous successive revisions as soon as they add a few words, which may result in a large number of revisions made by a single contributor (the coordinate points are those along the x-axis, on the right side of the scatterplots in Figure 9). Others contribute significant editions. For instance, the article *aéroportage* 'air transportation' in the French Wiktionary contains two senses ('transport by air' and 'airborne spread of a disease') with definitions and examples, a pronunciation, inflected forms, a synonym and a related word. The article was created in November 2020 in a one-shot edition and has not been modified since. Located at the (1, 1) coordinate point, it is undetectable. Future experiments will consist in taking into account the nature and length of contributions and possibly lead to a refinement of the method presented here.

I investigated above the presence of the OED Covid-related headwords in Wiktionary. A reverse question is: are Wiktionary's most heavily modified Covid-related headwords in the OED? The top-ranked ones are, except the various stigmatizing appellations *Wuhan virus, Wuhan pneumonia/flu* and *Chinese virus* that were used before the virus and resulting disease were officially named *Covid-19*. Humorous coinages such as *corona belly* or the derogatory *maskhole* may not be good candidates for OED inclusion. Other words such as *syndemic* and, maybe, *doomscrolling*, could be considered. Regarding semantic neology, the 2020 revisions of *antimask* in Wiktionary could draw attention to the possible need to update the OED article which only describes the grotesque dance.

4 Conclusion

The present study was based on the hypothesis that Wiktionary's most heavily modified articles can help detect new and existing headwords that are related to topical

22 Wiktionary's run-on entries were not taken into account in the present study.

events. Experimenting on the 2020 revisions and targeting Covid-related vocabulary proved successful and validated the hypothesis. One finding is that using only the number of unique contributors performs better than relying on the number of editions. In other words, Wiktionary's "crowd" of contributors is an asset for the task at hand. It does not mean that the number of revisions is not relevant. The conclusion to be drawn is rather that, given a set of articles having a similar number of revisions, the articles modified by the greatest diversity of contributors are the most likely to be related to topical events. Varying the ranking scores is also a good idea as it retrieves additional true positives.

Using Wiktionary's revision logs was considered a stopgap when no satisfactory diachronic corpus is available. When such a corpus exists, cross-checking the results of corpus-driven analyses and Wiktionary's history mining is certainly a good option.

A strength of the proposed method is that it is language and topic independent. Regarding language, the method is likely to perform well with the editions of Wiktionary that have the most active online communities. Regarding topics, one has to keep in mind that an event such as the Covid-19 pandemic is extraordinary, as were the two OED updates – and that *unprecedented* was the Oxford Languages word of the year 2020. Whether the suggested method is able to detect lexical innovations related to topical events that are less massive is an open question and the subject of future analyses. Trawling through the lists of candidates for the present study made me confident on that point. Other topics emerged, related for example to the US presidential election, identity and discrimination questions, police brutality – 2020 was also the year of the killing of George Floyd that brought the (pre-existing) Black Lives Matter movement to the forefront, with the BLM acronym ranking high in Wiktionary in the second trimester. Similar topics emerge from the French Wiktionary. In this language edition, a large number of feminine agent nouns were added. Though not precisely related to a timestamped event, and even though most of these nouns are feminized job titles related to forgotten professions, this trend is noteworthy.

Wiktionary revision logs give the opportunity to predict the past. A possible assessment of the suggested method consists in reiterating the experiments on the revisions of previous years and analysing what vocabulary emerges, related to which topic. In the meantime, the current study led me to examine the revision rate of *quarantine*, for which I observed a jump in 2020, and another back in 2009 (cf. Figure 16). Calculating the most frequently modified articles for that year enabled the detection of *swine flu*, which ranked first among the new articles (eclipsing the equally new *H1N1*), while, regarding existing headwords, *epidemic* ranked 111th globally and 16th in the second trimester; *quarantine* ranked 141st globally and 7th in the second trimester; and *mask* ranked 307th globally and 142nd in the third trimester. Hopefully, the future will allow for the detection of more enjoyable neologisms to be included in the dictionary. The present is apparently not a time for complacency. In the French Wiktionary, *vaxxie* 'a selfie taken while getting a Covid-19 vaccine', *centre de vaccination*

‘vaccination centre’, *Covid long* ‘long Covid’, *passeport vaccinal* ‘vaccine pass’ and *vaccino-sceptique* ‘sceptical about the usefulness or the efficiency of vaccines’ are among the most frequently modified neologisms during the first trimester of 2021 (with respective ranks of 4, 9, 17, 63 and 118). Regarding existing words, *vaccinodrome* ‘large capacity vaccination centre’ that was coined in 2020 and entered Wiktionary in March 2020 was not revised enough to be detected that year, but ranks 24th in the first semester of 2021, while *couvre-feu* ‘curfew’ ranks 15th in the second trimester. Which is a good point in favour of the suggested method, if not a light-hearted final note.

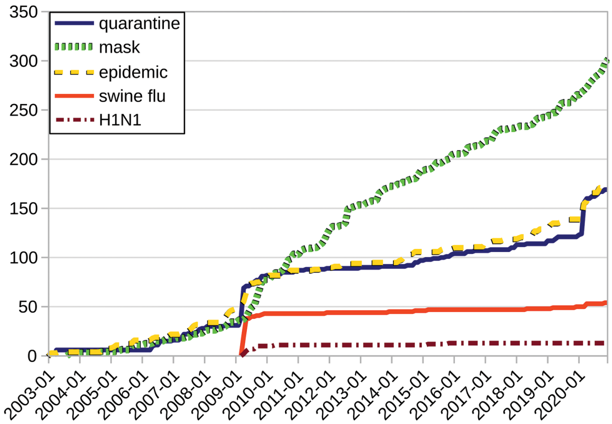


Figure 16: Neologisms and existing words showing notable increases in revisions in 2009.

Bibliography

Cook, Paul/Lau, Jey Han/Rundell, Michael/McCarthy, Diana/Baldwin, Timothy (2013): A lexicographic appraisal of an automatic approach for detecting new wordsenses. In: *Proceedings of the eLex 2013* conference, Tallinn, Estonia. Ljubljana: Trojina, Institute for Applied Slovene Studies; Tallinn: Eesti Keele Instituut, 49–98.

Fišer, Darja/Ljubešić, Nikola (2018): Distributional modelling for semantic shift detection. In: *International Journal of Lexicography* 32(2), 163–183.

Hamilton, William L./Leskovec, Jure/Jurafsky, Dan (2016): Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. In: *Proceedings of the Conference on Empirical Methods in Natural Language* Processing *(EMNLP’2016)*, Austin, Texas. Association for Computational Linguistics, 2116–2121.

Hanks, Patrick (2012): Corpus Evidence and Electronic Lexicography. In: Granger, Sylviane/Paquot, Magali (eds.): *Electronic Lexicography*. Oxford: Oxford University Press, 57–82.

Kilgarriff, Adam (1998): The hard parts of lexicography. *International Journal of Lexicography* 11(1), 51–54.

Kilgarriff, Adam (2009): Simple maths for keywords. In: Mahlberg, Michaela/González-Díaz, Victorina/Smith, Catherine (eds.): *Proceedings of Corpus Linguistics Conference*, Liverpool, UK. University of Liverpool.
Kutuzov, Andrey/Øvrelid, Lilja/Szymanski, Terrence/Velldal, Erik (2018): Diachronic word embeddings and semantic shifts: a survey. In: *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico. Association for Computational Linguistics, 1384–1397.
Landau, Sidney (2001): *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
Lih, Andrew (2004): Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. In: *Proceedings of the 5th International Symposium on Online Journalism*, Austin, Texas.
Renouf, Antoinette (2013): A finer definition of neology in English: The life-cycle of a word. In Hasselgård, Hilde/Ebeling, Jarle/Ebeling, Signe Oksefjell (eds.): *Corpus Perspectives on Patterns of Lexis*. Amsterdam/Philadalepia: John Benjamins, 177–208.
Rundell, Michael (2017): Dictionaries and crowdsourcing, wikis, and user-generated content. In: Hanks, Patrick/de Schryver, Gilles-Maurice (eds.): *International Handbook of Modern Lexis and Lexicography*. Berlin/Heidelberg: Springer, 1–16.
Sajous, Franck/Josselin-Leray, Amélie/Hathout, Nabil (2018): The Complementarity of Crowdsourced Dictionaries and Professional Dictionaries viewed through the Filter of Neology. In: *Lexis* 12.
Sajous, Franck/Calderone, Basilio/Hathout, Nabil (2020): ENGLAWI: From Human- to Machine-Readable Wiktionary. In: *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, France, 3016–3026.
Wolfer, Sascha/Müller-Spitzer, Carolin (2016): How many people constitute a crowd and what do they do? Quantitative analyses of revisions in the English and German Wiktionary editions. In: *Lexicos* 26, 347–371.